TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

Philipp Müller

Algorithms for Positioning with Nonlinear Measurement Models and Heavy-tailed and Asymmetric Distributed Additive Noise

Julkaisu 1395 • Publication 1395

Tampere 2016

Tampereen teknillinen yliopisto. Julkaisu 1395
Tampere University of Technology. Publication 1395

Philipp Müller

Algorithms for Positioning with Nonlinear Measurement Models and Heavy-tailed and Asymmetric Distributed Additive Noise

Thesis for the degree of Doctor of Science in Technology to be presented with due permission for public examination and criticism in Sähkötalo Building, Auditorium S2, at Tampere University of Technology, on the 26th of August 2016, at 12 noon.

Tampereen teknillinen yliopisto - Tampere University of Technology
Tampere 2016

ISBN 978-952-15-3769-1 (printed)
ISBN 978-952-15-3784-4 (PDF)
ISSN 1459-2045

Abstract

Determining the unknown position of a user equipment using measurements obtained from transmitters with known locations generally results in a nonlinear measurement function. The measurement errors can have a heavy-tailed and/ or skewed distribution, and the likelihood function can be multimodal.

A positioning problem with a nonlinear measurement function is often solved by a nonlinear least squares (NLS) method or, when filtering is desired, by an extended Kalman filter (EKF). However, these methods are unable to capture multiple peaks of the likelihood function and do not address heavy-tailedness or skewness. Approximating the likelihood by a Gaussian mixture (GM) and using a GM filter (GMF) solves the problem. The drawback is that the approximation requires a large number of components in the GM for a precise approximation, which makes it unsuitable for real-time positioning on small mobile devices.

This thesis studies a generalised version of Gaussian mixtures, which is called GGM, to capture multiple peaks. It relaxes the GM's restriction to non-negative component weights. The analysis shows that the GGM allows a significant reduction of the number of required Gaussian components when applied for approximating the measurement likelihood of a transmitter with an isotropic antenna, compared with the GM. Therefore, the GGM facilitates real-time positioning in small mobile devices. In tests for a cellular telephone network and for an ultra-wideband network the GGM and its filter provide significantly better positioning accuracy than the NLS and the EKF.

For positioning with nonlinear measurement models, and heavy-tailed and skewed distributed measurement errors, an Expectation Maximisation (EM) algorithm is studied. The EM algorithm is compared with a standard NLS algorithm in simulations and tests with realistic emulated data from a long term evolution network. The EM algorithm is more robust to measurement outliers. If the errors in

training and positioning data are similar distributed, then the EM algorithm yields significantly better position estimates than the NLS method. The improvement in accuracy and precision comes at the cost of moderately higher computational demand and higher vulnerability to changing patterns in the error distribution (of training and positioning data). This vulnerability is caused by the fact that the skew-t distribution (used in EM) has 4 parameters while the normal distribution (used in NLS) has only 2. Hence the skew-t yields a closer fit than the normal distribution of the pattern in the training data. However, on the downside if patterns in training and positioning data vary than the skew-t fit is not necessarily a better fit than the normal fit, which weakens the EM algorithm's positioning accuracy and precision. This concept of reduced generalisability due to overfitting is a basic rule of machine learning.

This thesis additionally shows how parameters of heavy-tailed and skewed error distributions can be fitted to training data. It furthermore gives an overview on other parametric methods for solving the positioning method, how training data is handled and summarised for them, how positioning is done by them, and how they compare with nonparametric methods. These methods are analysed by extensive tests in a wireless area network, which shows the strength and weaknesses of each method.

Preface

The work presented in this thesis was carried out at the Positioning Algorithms group at the Department of Mathematics and at the Department of Automation Science and Engineering, Tampere University of Technology between 2011 and 2016. The research was funded by the Doctoral Programme of Tampere University of Technology's President and Nokia Corporation. I gratefully acknowledge the additional financial support from the Finnish Doctoral Programme in Computational Sciences (FICS) and TUT-Foundation (TTY-ssätiö).

I am grateful to my supervisor Prof. Robert Piché for guiding and encouraging me during my doctoral studies. He has been always available for discussions about my research and organisational aspects. I thank Dr. Simo Ali-Löytty for counselling me during my studies and for his valuable suggestions how to improve my thesis. I thank my colleagues from the Positioning Algorithms group, especially Henri Nurminen and Juha Ala-Luhtala for our countless helpful discussions and joined work. I am grateful to Dr. Matti Raitoharju for his work on our common publications. I thank Dr. Lauri Wirola and Dr. Jari Syrijänne at Nokia/ here for providing me an insight into industrial research.

I am grateful to Asoc. Prof. Henk Wymeersch for hosting me in the autumn of 2012, and our fruitful discussions during my research visit to Chalmers University of Technology. I thank Prof. Gonzalo Seco-Granados for hosting me in the autumn of 2014 at Universitat Autònoma de Barcelona, Spain, and Dr. José Antonio Del Peral-Rosado for our discussions and his work on our joined publication. I also want to express my gratitude to the pre-examiners of my thesis, Dr. Pau Closas and Asoc. Prof. Gustaf Hendeby, for their comments

and suggestions. I thank Prof. Giuseppe Abreu for acting as opponent in the public examination of my thesis.

I thank my family for their understanding and for supporting my decision to pursue my doctoral studies more than 1 300 kilometres away from them. I express my gratitude to my friends for their support and friendship, especially to André, Matti, Jaakko and Liang. Finally, I am deeply grateful to my girlfriend Shengye for her love and her support.

Tampere, May 2016,

Philipp Müller

Contents

List of publications

This thesis consist of an introduction, and the following publications and one unpublished manuscript.

PUBLICATIONS

P1. Philipp Müller, Matti Raitoharju, Simo Ali-Löytty, Laura Wirola, and Robert Piché. "A Survey of Fingerprinting and Parametric Fingerprint-Positioning Methods". In *Gyroscopy and Navigation*, Vol. 7, No. 2, pages 107–127, April 2016.

P2. Philipp Müller, Simo Ali-Löytty, Marzieh Dashti, Henri Nurminen, and Robert Piché. "Gaussian mixture filter allowing negative weights and its application to positioning using signal strength measurements". In *Proceedings of the 2012 9th Workshop on Positioning, Navigation and Communication (WPNC)*, pages 71–76, Dresden, Germany, March 2012.

P3. Philipp Müller, Henk Wymeersch, and Robert Piché. "UWB positioning with generalized Gaussian mixture filters". *IEEE Transactions on Mobile Computing*, Vol. 13, No. 10, pages 2406–2414, October 2014.

P4. Philipp Müller and Robert Piché. "Statistical trilateration with skew-t errors". In *2015 International Conference on Localization and GNSS (ICL-GNSS)*, 6 pages, Gothenburg, Sweden, June 2015.

UNPUBLISHED MANUSCRIPT

R1. Philipp Müller, José A. del Peral-Rosado, Robert Piché, and Gonzalo Seco-Granados. Statistical Trilateration with Skew-t Distributed Errors in LTE networks.

The main contributions of the publications and the unpublished manuscript are

P1. The paper provides a unified, comprehensive overview on parametric fingerprint-positioning methods with detailed explanation of their offline and online phases. It defines what are parametric methods, which features they share with nonparametric methods and what are the differences between the groups. Several of the parametric methods proposed in the literature

together with one nonparametric method are compared by tests in the same wireless local area network for different scenarios. The experimental results uncover the advantages and disadvantages of the different methods.

P2. A generalisation of the Gaussian mixture approach, which allows negative component weights, is developed. The approach is called generalised Gaussian mixture (GGM; in the paper it is denoted as GMA). It is a computational light and efficient method for positioning scenarios with significant nonlinear measurement models. The GGM provides a better approximation of the measurement likelihood than the extended Kalman filter (EKF), which underestimates the posterior covariance.

P3. The GGM from [P2] for positioning is analysed in terms of computational complexity and approximation quality. The method for determining the Gaussian components' parameters is modified; and the GGM filter's positioning performance is evaluated for real data from an ultra-wideband (UWB) network and compared with an EKF's performance. In the analysis part it is shown that the GGM's approximation quality for the measurement likelihood depends on the measurement's uncertainty. The tests show that the GGM filter outperforms the EKF in different scenarios and it provides especially in the beginning of the filtered track superior performance compared with the EKF.

P4. An Expectation Maximisation (EM) algorithm for solving statistical trilateration problems with skew-t distributed measurement errors is introduced and derived. The EM is a stable method that converges when searching for a maximum a-priori (MAP) position estimate and is fast compared with Markov-chain Monte-Carlo (MCMC) methods. It handles heavy-tailedness and/ or skewness in measurement error/ noise distributions, i.e. two types of non-Gaussianity that are not handled well by nonlinear least squares (NLS) methods. The EM and the descending Gauss-Newton (dGN) already existed in statistical literature; the paper's contribution is their application to the positioning framework.

R1. The EM algorithm from [P4] is analysed in detail using realistic emulated data from an LTE network. Its positioning accuracy is compared with that of a standard NLS algorithm under ideal

simulated conditions and under unfavourable real-world conditions. The analysis shows that the EM improves positioning accuracy when measurement errors follow a heavy-tailed distribution even with mild skewness, but requires high-quality training data to do so.

My role in the shared publications and the unpublished manuscript:

P1. I wrote the manuscript.

P2. I developed, implemented and tested the generalised Gaussian mixture approach based on the idea proposed by the co-author Simo Ali-Löytty. I implemented the algorithms, carried out the simulations, and wrote the manuscript except for section III.

P3. I implemented the algorithms, organised and executed the measurement collection campaign, run all tests and wrote the manuscript.

P4. I derived and implemented the EM and the Gibbs sampler algorithm proposed by the co-author Robert Piché, implemented the standard NLS algorithm, carried out the simulations, and wrote the manuscript.

R1. I implemented the algorithm, carried out the tests and wrote the manuscript, except for parts of the introduction and subsections II.D, III.B and IV.A.

Abbreviations

2G	second generation
3G	third generation
AP	access point
BFGS	Broyden-Fletcher-Goldfarb-Shanno
BS	base station
BS-ID	base station identifier
CA	coverage area
CAF	coverage area filter
CDF	cumulative distribution function
CKF	cubature Kalman filter
dGN	descending Gauss-Newton
DOF	degrees-of-freedom
DR	dead reckoning
E-step	expectation step
EKF	extended Kalman filter
EM	Expectation Maximisation
EPA5	Extended Pedestrian A model 5
ETU70	Extended Typical Urban model 70
FP	fingerprint
GGM	generalised Gaussian mixture
GGMF	generalised Gaussian mixture filter
GHF	Gauss-Hermite filter
GHKF	Gauss-Hermite Kalman filter
GM	Gaussian mixture
GMF	Gaussian mixture filter
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GS	Gibbs sampler
GSM	Global System for Mobile Communications
ID	identifier
IMU	inertial measurement unit
INS	inertial navigation system
IRLS	Iterative Reweighted Least Squares
KLD	Kullback-Leibler divergence
LBS	location-based services
LOS	line of sight
LTE	Long Term Evolution
M-step	maximisation step

MAP	maximum a-posteriori
MCMC	Markov-chain Monte-Carlo
ML	maximum likelihood
MQKF	multiple quadrature Kalman filter
NLOS	non-line of sight
NLS	nonlinear least squares
NN	nearest neighbour
PDF	probability density function
PDR	pedestrian dead reckoning
PL	path loss
PRS	positioning reference signal
QKF	quadrature Kalman filter
RM	radio map
RSS	received signal strength
RTOF	roundtrip time of flight
SCKF	square-root cubature Kalman filter
SLAM	simultaneous localisation and mapping
SNR	signal-to-noise ratio
SPD	symmetric positive definite
SGQF	sparse-grid quadrature filter
SQKF	square-root quadrature Kalman filter
SSM	state space model
TDOA	time difference of arrival
TOA	time of arrival
UE	user equipment
UKF	unscented Kalman filter
UT	unscented transformation
UWB	ultra-wideband
VB	variational Bayes
WKNN	weighted k-nearest neighbours
WLAN	wireless local area network

Symbols

A	apparent transmission power
α	scaling parameter for dGN
α_j	weight of Gaussian component
b_j	cell identifier
β	configuration parameter for GGM
$\mathbf{C}_t$	cross-covariance of state measurement
$\bar{c}$	weight in GGM with 2 components
$\mathbf{c}_k$	reference node location
d	distance between transmitter and receiver
d_{dGN}	position update dGN
d_{EM}	position update EM
$\mathbf{e}_{K\times 1}$	all-ones vector
ϵ	minimum standard deviation for GGM
ϵ_{dGN}	minimum position update dGN
ϵ_{EM}	minimum position update EM
$\text{erfc}(\cdot)$	complementary error function
$\text{erfcx}(\cdot)$	scaled complementary error function
η	path loss exponent
$f(\cdot)$	cost function for dGN
$\mathbf{f}_{t-1}(\cdot)$	state transition function
$\mathbf{F}_t$	state transition matrix
$\Gamma(\cdot)$	gamma function
$\text{Gamma}(s,\theta)$	Gamma distribution
$\mathbf{h}_t(\cdot)$	measurement function
$\mathbf{H}_t$	measurement transition matrix
$\text{HN}\left(\mu,\sigma^2\right)$	half-normal distribution
ID_k	transmitter identifier
$\mathbf{I}_{K\times K}$	identity matrix
$\mathbf{K}_t$	filter gain
κ	parameter for UKF
$\mathbf{l}_{\text{FP}_j}$	location of fingerprint FP_j
μ	mean value
m	mean of normal fitted to range errors
$\mathbf{m}$	prior position estimate
$\text{MVN}\left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right)$	multivariate normal distribution
$\text{MVN}_{\boldsymbol{\Sigma}}^{\boldsymbol{\mu}}(\mathbf{x})$	multivariate normal distribution
n_{dGN}	number of iterations in dGN
n_{EM}	number of iterations in EM

$n_{\mathbf{x}}$	dimension of vector $\mathbf{x}$
$\mathrm{N}(\mu, \sigma^2)$	normal distribution
ω_j	weight of fingerprint location
$\mathbf{P}$	prior position estimate's covariance
$\mathbf{P}_t$	covariance of state at t
$\mathbf{P}_t^-$	mean of prior covariance
P_{RSS}	received signal strength
$\mathbf{P}_{\mathrm{RSS}}$	vector of received signal strength values
$p_{\underline{\mathbf{x}}}(\mathbf{x})$	PDF of random vector $\mathbf{x}$
$p(\mathbf{x})$	PDF of random vector $\mathbf{x}$
$p_{\underline{\mathbf{x}}\|\underline{\mathbf{y}}}(\mathbf{x}\|\mathbf{y})$	conditional PDF of random vector $\underline{\mathbf{x}}$ given $\mathbf{y}$
$p(\underline{\mathbf{x}}\|\mathbf{y})$	conditional PDF of random vector $\underline{\mathbf{x}}$ given $\mathbf{y}$
$\phi(\cdot)$	PDF standardised Normal distribution
$\Phi(\cdot)$	CDF standardised Normal distribution
$\mathbf{Q}_{t-1}$	state transition error covariance
$\mathbf{R}_t$	measurement error covariance
$\mathbb{R}^n$	set of n-dimensional real numbered vectors
$\mathbf{S}_t$	covariance of measurement
$\mathbf{\Sigma}$	covariance matrix
$\|\mathbf{\Sigma}\|$	determinant of matrix $\mathbf{\Sigma}$
$\sigma_{\max}$	standard deviation first component GGM
$\sigma_{\min}$	standard deviation second component GGM
σ_w^2	shadow fading variance
$\mathrm{ST}\left(\xi, \sigma^2, \lambda, \nu\right)$	skew-t distribution
T_0	number of burn-in samples
T	number of retained samples
$\mathrm{T}(\xi, \sigma^2, \nu)$	t-distribution
$t_\nu(\cdot)$	PDF standardised t-distribution
$T_\nu(\cdot)$	CDF standardised t-distribution
t_k	hyperparameter hierarchical model of skew-t
τ_k	hyperparameter hierarchical model of skew-t
$\boldsymbol{\theta}$	ellipse parameters
$\mathrm{TN}\left(\mu, \sigma^2\right)$	truncated normal distribution
u_{dGN}	update position estimate dGN
u_{EM}	update position estimate EM
$\mathrm{U}(a, b)$	uniform distribution
v	measurement error
$\mathbf{v}$	vector of ranging errors
$\underline{\mathbf{v}}_t$	measurement noise

$\underline{\mathbf{w}}_{t-1}$	state transition noise
$\underline{x}$	random variable
$\underline{\mathbf{x}}$	random vector
$\mathbf{x}$	column vector
$\mathbf{X}$	matrix
$\mathbf{x}_0$	initial state
$\hat{\mathbf{x}}_t$	mean of state at t
$\hat{\mathbf{x}}_t^-$	mean of prior state
$\hat{\mathbf{x}}$	MAP position estimate
$\hat{\mathbf{x}}_{\mathrm{pm}}$	posterior mean position estimate
$\mathbf{x}^{(j)}$	jth sample for GS
$\mathcal{X}_{t-1}^{\cdot(i)}$	sigma point (prediction step)
$\hat{\mathcal{X}}_t^{\cdot(i)}$	propagated sigma point (prediction step)
$\mathcal{X}_t^{\cdot-(i)}$	sigma point (update step)
$\hat{\mathcal{Y}}_t^{(i)}$	propagated sigma point (update step)
y_k	ranging measurement
$\tilde{y}_k$	modified ranging measurement
$\bar{y}_k$	modified ranging measurement
$\mathbf{z}_n$	fingerprint location
ζ	parameter for UKF

CHAPTER 1

Introduction

This thesis consists of an introduction, three articles published in scientific conferences and journals, one article accepted for publication in a scientific journal and one unpublished manuscript. The purpose of this introductory chapter is to present a short unified background for the publications [P1]–[P4] and the unpublished manuscript [R1], and summarise their results and put their contributions into context.

The chapter is divided into four sections. Section 1 presents an introduction to positioning and explains which questions from that research area my dissertation wants to answer. In Section 2 Bayes' rule, Bayesian filtering and statistical trilateration are explained briefly. The positioning methods presented in [P1]–[P4] and [R1] can be divided in two parts; an offline phase, in which a radio map is generated or parameters of distributions are fitted to training data, and an online phase, in which the radio map and/ or the fitted parameters are used for positioning. Section 3 summarises how radio maps are generated for the proposed methods and the methods used for comparison in [P1], [P2], and [P3]. In Section 4, it is explained how parameters of normal and skew-t distributions are fitted to data, which is needed in [P4] and [R1]. In Section 5 the positioning algorithms developed in [P2] and [P4], which are analysed in [P3] and

[R1], are briefly introduced and their main features are explained. Furthermore, the reference methods used in [P1]–[P4] and [R1] are explained.

The contributions of my publications are presented in the different sections as follows: The generalised Gaussian mixture (GGM) from [P2], [P3] is defined and explained in Subsection 5.3. This section also summarises the results of the simulations and tests with real-world data in [P2] and [P3]. The Expectation Maximisation (EM) algorithm from [P4], [R1] is presented in Subsection 5.5. This section also summarises the results of the numerical experiment and the tests with realistic emulated data in [P4] and [R1]. The Gibbs sampler for fitting parameters from [P4], [R1] is presented in Section 4. Its possible application in the online phase is described in Subsection 5.6. The content of publication [P1] can be found in both Section 3 and Section 5.

1 Background

Positioning techniques have become the backbones of location-aware applications for use in commercial, public service and military networks [32, 34, 55]. Those applications are manifold. They are used for vehicle navigation, intelligent transport systems, vehicle tracking and fleet management. Applications for public services include emergency services (e.g. E911 in North America, E112 in Europe), medical services (e.g. patient and equipment surveillance in hospitals), and rescue operations (e.g. locating fire fighters in burning buildings).

In recent years a tremendous amount of location-based services (LBS), intended for usage on smartphones, have been developed that require knowledge of the phone's location. The reasons for the success of those so-called second generation LBS are manifold [16]. Unlike the first generation LBS, the users decide whether they use a LBS, and reveal their location, thus allowing them to preserve their privacy. Furthermore, new LBS are application oriented and proactive, meaning that the offered services depend on the user's profile rather than their location and are launched automatically if certain conditions are fulfilled. Another reason for the significant increase in available LBS is that nowadays almost everybody can produce LBS, while in earlier days the mobile phone network operators were the only providers. For a more extensive discussion the reader is referred to [16].

LBS include local search, suggesting local points of interest, geotagging of photos and videos, location sensitive billing, and targeted advertising [24, 34, 55, 59, 78, 81, 85]. Running them on small mobile devices limits the energy, memory, bandwidth, and computational resources that positioning algorithms are allowed to use [55].

For outdoor environments positioning relies mostly on Global Navigation Satellite System (GNSS) signals. Because GNSS receivers are nowadays relatively inexpensive, most mobile user equipments (UEs) such as smartphones contain them, allowing ubiquitous positioning outdoors. However, using GNSS receivers for positioning has several drawbacks. Firstly, the receivers use a significant amount of energy. For example, Constandache et al. [22] found that their mobile phone's

net battery life was 9 hours when the Global Positioning System (GPS) receiver was used, while it was 40 and 60 hours with only WiFi receiver or Global System for Mobile Communications (GSM) respectively used. Secondly, in indoor environments as well as forest canopies and urban canyons GNSS-based positioning yields unreliable location estimates or fails completely due to poor signal penetration by GNSS signals in those environments. Therefore, alternative measurements have to be used for positioning in these environments. In this thesis the focus is on measurements from radio networks.

The probably most widely used radio network type are cellular telephone networks. In these networks signals can be send from the UE to the base station (BS) and vice versa [61, p. 10 f.]. Older cellular networks such as second generation (2G) and third generation (3G) were not developed specifically for positioning. However, positioning algorithms that use information such as cell identifiers (IDs) and received signal strength (RSS) values can achieve sufficient accuracy for applications such as local search or weather forecast [87]. In newer cellular networks signals that are designed for positioning are specified. For example, in release 9 of the Long Term Evolution (LTE) standard, the positioning reference signal (PRS) is introduced, which can achieve positioning accuracy on the order of 10 m [27]. In [2] trilateration techniques for estimating the UE's position in an LTE network are specified. These techniques use measurements of distances (ranges) between BSs with known location and the UE, which are derived from time-delay measurements.

Another type of radio networks widely used for positioning is a wireless local area network (WLAN). Similarly to 2G and 3G, WLANs were not designed for positioning, but research has shown that WLAN-based positioning can achieve high positioning accuracy and therefore often is preferred over cellular-based positioning [68]. The main reasons for the higher accuracy are the generally higher BS density in WLANs (BS is usually called access point (AP) in WLANs) compared with cellular networks and the smaller area in which an AP's signal can be received (this area is called coverage area). Most techniques for positioning in WLANs use RSS readings, which are correlated with the UE location [61, p. 47], and can be applied also for positioning in cellular networks. In [P1] a wide selection of such algorithms is

analysed and their positioning performance is evaluated in extensive tests with real-world WLAN data.

Ultra-wideband (UWB) is another radio network type that has been considered for positioning [3, 24, 32, 43]. As the name suggests it uses extremely large bandwidths, which enables accurate range estimates and high reliability. Similarly to LTE networks, range estimates are derived from time-delay measurements.

There exist various other systems, such as Bluetooth, wireless sensor networks and inertial navigation systems (INS) that are being used for positioning. However, they are out of scope for this thesis.

Similarly there exist various methods for positioning in the various network types. This thesis focuses on statistical trilateration and Bayesian filtering. In [25] a detailed overview on different positioning techniques is given. Those include methods for distributed and cooperative positioning, fingerprinting, simultaneous localisation and mapping (SLAM), and fusion. In distributed cooperative positioning techniques receivers with unknown position share information to gain (further) knowledge about their positions. Fingerprinting is the process of storing features observable in certain locations in a database in the so-called offline phase, and then inferring the UE position by comparing the features the UE observes in its unknown location with the database entries. In SLAM the mobile device simultaneously estimates its location and the map of its environment. Fusion techniques combine the information from several sources to improve positioning performance. Examples are the use of GNSS receivers, map information and an inertial measurement unit (IMU) for positioning in outdoor environments and the use of WLAN receiver, floor plan information and an IMU in indoor environments. More detail can be found, for example, in [25].

Positioning in radio networks and its requirements

For determining the two-dimensional UE location by trilateration unambiguously the distances to at least three BSs with known locations have to be known. The measurements are often either RSS or time-delay measurements, from which ranges are derived. Thus, the measurement function that describes the relation between the range

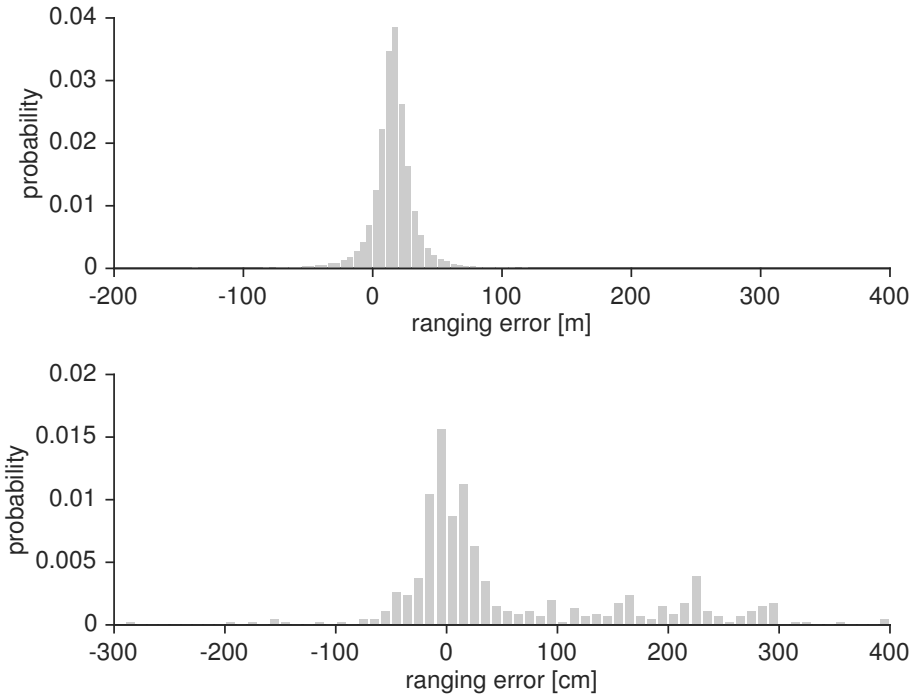


Figure 1: Histograms of ranging errors from LTE's EPA5 channel model (upper plot) and UWB NLOS ranging error data from [P3].

measurements and the UE location is often nonlinear.

The measurements, in general, contain errors, corrupting the observed signal of interest. Therefore, statistical measurement models (aka measurement equations) that describe the relation between range measurements, UE location and measurement errors are widely used. If the UE has to be tracked then the measurement model is extended by a state transition model that describes how the UE location changes over time and accounts for errors as well. For tracking, Bayesian filtering can be employed.

Measurement errors are often modelled as zero-mean normal distributed random variables for the sake of simplicity or to make the problem analytically tractable. However, this assumption is not always a faithful representation of the actual distribution. Figure 1 shows histograms for ranging errors from an LTE network (upper plot) and from an UWB network, for which the assumption is clearly violated. The LTE range errors are from the Extended Pedestrian A

model 5 (EPA5) channel model. The UWB range errors are errors in the non-line of sight (NLOS) range measurements from [P3]. The histograms show that ranging errors can be asymmetric distributed (skewed) and contain a significant amount of outliers (heavy-tailed aka high-kurtosis). A similar behaviour was detected in [47].

The ideal positioning algorithm uses a measurement equation that models the measurements and their errors accurately and provides precise position estimates. At the same time, however, it has to enable real-time positioning on small mobile devices such as smartphones. Thus, it has only limited energy resources and computational power at hand, meaning that the algorithm should be of relatively low computational demand/ complexity. Furthermore, it should be easy to implement and maintain to keep the costs for manufacturers and providers of the positioning service low.

Research questions

Keeping all these criteria in mind a compromise has to be found, which leads to the two research questions that are considered in this thesis:

1. How can the UE position be calculated precisely, reliably, efficiently and with low cost for nonlinear measurement models?

2. How can heavy tailed and/ or skewed distributed measurement errors/ noise be properly accounted for in positioning problems?

If the measurement model is mildly nonlinear then the answer to the first question could be the use of a nonlinear least squares (NLS) method for static positioning or an extended Kalman filter (EKF) for tracking [15, 40]. However, for RSS or time-delay based ranges the measurement function is highly, i.e. significantly, nonlinear. Under these circumstance the EKF suffers from several drawbacks. Firstly, for highly nonlinear models the EKF can seriously underestimate the posterior covariance [8]. Secondly, the EKF is unable to capture multiple peaks of the likelihood function, which occurs often for nonlinear measurement models. The second drawback also holds for NLS algorithms.

Particle filters achieve high accuracy in highly nonlinear conditions [13] but are computationally demanding, making real-time positioning on small mobile devices difficult, if not impossible. Another approach that enables capturing multiple peaks is to approximate the likelihood function by a Gaussian mixture (GM) and to use a GM filter (GMF) for tracking.

In order to keep the computational resources used by the GMF on a level that can be handled on small mobile devices the number of Gaussian components has to be kept small without losing significant information. The GGM introduced in [P2] and analysed thoroughly in [P3] achieves both goals by relaxing the GM's non-negativity constraint on component weights.

The third drawback of the EKF, which also holds for NLS algorithms and many other techniques, is that it models errors as zero-mean normal distributed. A distribution that captures the heavy-tailedness and skewness seen in Figure 1 is the skew-t distribution. In [P4] an EM algorithm for determining the maximum a-posteriori (MAP) estimate for the UE position under the assumption of skew-t distributed range errors is developed. The algorithm is thoroughly analysed in [R1] in an LTE network using the EPA5 channel, which represents pedestrian walking and open environments, and the Extended Typical Urban model 70 (ETU70) channel, which represents urban environments with high mobility.

Alternative approaches for solving the positioning problem under the assumption of skew-t distributed range errors are maximum likelihood (ML) estimation and Markov-chain Monte-Carlo (MCMC) techniques. However, MAP estimation has the advantage over ML estimation that it allows the use of a non-uniform prior. MCMC methods provide credibility intervals for the position estimate, unlike MAP estimation, but are too computationally demanding for execution on small mobile devices.

2 Mathematical background

In this thesis, in [P1]–[P3] both static and filtered positioning problems are considered; in [P4] and [R1] static positioning problems are considered. This section summarises the estimation theory applied for solving the positioning problems. It furthermore summarises some of the alternative methods used in the literature.

2.1 Bayes' rule and Bayesian filtering

Bayes' rule is used widely in this thesis for estimating parameters of distributions and for solving static positioning problems.

The aim of both tasks is to find an estimate for the $n_{\mathbf{x}_t}$-dimensional state $\mathbf{x}_t \in \mathbb{R}^{n_{\mathbf{x}_t}}$ given $n_{\mathbf{y}_t}$ scalar measurements $\mathbf{y}_t \in \mathbb{R}^{n_{\mathbf{y}_t}}$ and a prior $p_{\underline{\mathbf{x}}_t}(\mathbf{x}_t)$. The prior models information about the state before any measurements have been observed. If such information is unavailable then a non-informative prior can be used [74, p. 19]. The posterior probability density function (PDF) of the state can be computed by applying Bayes' rule (see e.g. [74, p. 19]) as

$$p_{\underline{\mathbf{x}}_t|\underline{\mathbf{y}}_t}(\mathbf{x}_t|\mathbf{y}_t) \propto p_{\underline{\mathbf{y}}_t|\underline{\mathbf{x}}_t}(\mathbf{y}_t|\mathbf{x}_t) p_{\underline{\mathbf{x}}_t}(\mathbf{x}_t). \tag{1}$$

For tracking problems the Bayesian filtering framework is used. In the publications the additive noise, discrete-time state space model (SSM)

$$\underline{\mathbf{x}}_t = \mathbf{f}_{t-1}(\underline{\mathbf{x}}_{t-1}) + \underline{\mathbf{w}}_{t-1}, \tag{2a}$$
$$\underline{\mathbf{y}}_t = \mathbf{h}_t(\underline{\mathbf{x}}_t) + \underline{\mathbf{v}}_t \tag{2b}$$

and some of its special cases are used. The functions $\mathbf{f}_{t-1}(\cdot)$ (state transition function) and $\mathbf{h}_t(\cdot)$ (measurement function) are assumed to be known and are possibly nonlinear. The subscript t is a time index, which takes values $t = 1, 2, \ldots$ The additive noises (aka errors) $\underline{\mathbf{w}}_{t-1}$ and $\underline{\mathbf{v}}_t$ are assumed to be white, mutually independent and independent of the initial state $\mathbf{x}_0$. Their PDFs are denoted $p_{\underline{\mathbf{w}}_{t-1}}$ and $p_{\underline{\mathbf{v}}_t}$ hereafter.

The aim of filtering is to find the marginal conditional probability density function $p_{\underline{\mathbf{x}}_t|\underline{\mathbf{y}}_{1:t}}(\mathbf{x}_t|\mathbf{y}_{1:t})$ of state $\underline{\mathbf{x}}_t$ given the observations

$\mathbf{y}_{1:t} \triangleq (\mathbf{y}_1, \ldots, \mathbf{y}_t)$ up to time t (marginal posterior). This can be done recursively using the following relations [40].
Prediction:

$$p_{\underline{\mathbf{x}}_t|\underline{\mathbf{y}}_{1:t-1}}(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p_{\underline{\mathbf{x}}_t|\underline{\mathbf{x}}_{t-1}}(\mathbf{x}_t|\mathbf{x}_{t-1}) p_{\underline{\mathbf{x}}_{t-1}|\underline{\mathbf{y}}_{1:t-1}}(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) \,\mathrm{d}\mathbf{x}_{t-1}; \quad (3)$$

Update:

$$p_{\underline{\mathbf{x}}_t|\underline{\mathbf{y}}_{1:t}}(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{p_{\underline{\mathbf{y}}_t|\underline{\mathbf{x}}_t}(\mathbf{y}_t|\mathbf{x}_t) p_{\underline{\mathbf{x}}_t|\underline{\mathbf{y}}_{1:t-1}}(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{\int p_{\underline{\mathbf{y}}_t|\underline{\mathbf{x}}_t}(\mathbf{y}_t|\mathbf{x}_t) p_{\underline{\mathbf{x}}_t|\underline{\mathbf{y}}_{1:t-1}}(\mathbf{x}_t|\mathbf{y}_{1:t-1}) \,\mathrm{d}\mathbf{x}_k}, \quad (4)$$

where the transition PDF is $p_{\underline{\mathbf{x}}_t|\underline{\mathbf{x}}_{t-1}}(\mathbf{x}_t|\mathbf{x}_{t-1}) = p_{\underline{\mathbf{w}}_{t-1}}(\mathbf{x}_t - \mathbf{f}(\mathbf{x}_{t-1}))$ and the likelihood is

$$p_{\underline{\mathbf{y}}_t|\underline{\mathbf{x}}_t}(\mathbf{y}_t|\mathbf{x}_t) = p_{\mathbf{v}_t}(\mathbf{y}_t - \mathbf{h}(\mathbf{x}_t)). \quad (5)$$

The PDF of the initial state $p_{\underline{\mathbf{x}}_0|\underline{\mathbf{y}}_{1:0}}(\mathbf{x}_0|\mathbf{y}_{1:0}) = p_{\underline{\mathbf{x}}_0}(\mathbf{x}_0)$ is used as initial condition for the recursion. Knowledge of the posterior (4) enables computing a state estimate that is optimal with respect to a given criterion.

For the cases analysed in [P1]–[P3] it is impossible to determine the conditional PDF analytically, and approximative methods for computing the posterior mean are used instead.

Extended Kalman filter

One approximative approach that is used is the extended Kalman filter, which uses Kalman filtering and applies it to a locally linearised version of system (2). The EKF assumes noises $\underline{\mathbf{w}}_{t-1}$ and $\underline{\mathbf{v}}_t$ to be zero-mean multivariate normal distributed, i.e. $\underline{\mathbf{w}}_{t-1} \sim \mathrm{MVN}(\mathbf{0}, \mathbf{Q}_{t-1})$ and $\underline{\mathbf{v}}_t \sim \mathrm{MVN}(\mathbf{0}, \mathbf{R}_t)$. A derivation of the equations presented hereafter can be found, for example, in [74, p. 69 ff.].

Given a multivariate normal approximation of the posterior at "time" $t-1$, with mean $\hat{\mathbf{x}}_{t-1}$ and covariance $\mathbf{P}_{t-1}$, the prediction (3) is solved using a linearised state (aka dynamic) equation by

$$\hat{\mathbf{x}}_t^- = \mathbf{f}_{t-1}(\hat{\mathbf{x}}_{t-1}), \quad (6a)$$

$$\mathbf{P}_t^- = \mathbf{F}_t \mathbf{P}_{t-1} \mathbf{F}_t^T + \mathbf{Q}_{t-1}, \quad (6b)$$

where $\hat{\mathbf{x}}_t^-$ is mean and $\mathbf{P}_t^-$ is covariance of a multivariate normal distribution, and $\mathbf{F}_t$ is obtained from the linearisation of the state model around $\hat{\mathbf{x}}_{t-1}$:

$$\mathbf{F}_t = \left.\frac{\partial \mathbf{f}_{t-1}(\mathbf{x}_t)}{\partial \mathbf{x}_t}\right|_{\hat{\mathbf{x}}_{t-1}}.$$

This means that $\underline{\mathbf{x}}_t \left| \left(\underline{\mathbf{y}}_{1:t-1} = \mathbf{y}_{1:t-1} \right) \sim \mathrm{MVN}\left(\hat{\mathbf{x}}_t^-, \mathbf{P}_t^-\right)\right.$.

In the update step a multivariate normal distribution with posterior mean $\hat{\mathbf{x}}_t$ and posterior covariance matrix $\mathbf{P}_t$, is computed by

$$\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_t^- + \mathbf{K}_t\left(\mathbf{y}_t - \mathbf{h}(\hat{\mathbf{x}}_t^-)\right), \tag{7a}$$

$$\mathbf{P}_t = \mathbf{P}_t^- - \mathbf{K}_t\mathbf{H}_t\mathbf{P}_t^-, \tag{7b}$$

where the Kalman gain $\mathbf{K}_t$ is

$$\mathbf{K}_t = \mathbf{P}_t^-\mathbf{H}_t\left(\mathbf{H}_t\mathbf{P}_t^-\mathbf{H}_t^T + \mathbf{R}_t\right)^{-1}$$

and $\mathbf{H}_t$ is obtained from the linearisation of the measurement model around $\hat{\mathbf{x}}_t^-$:

$$\mathbf{H}_t = \left.\frac{\partial \mathbf{h}_t(\mathbf{x}_t)}{\partial \mathbf{x}_t}\right|_{\hat{\mathbf{x}}_t^-}.$$

This means that $\underline{\mathbf{x}}_t \left| \left(\underline{\mathbf{y}}_{1:t} = \mathbf{y}_{1:t} \right) \sim \mathrm{MVN}(\hat{\mathbf{x}}_t, \mathbf{P}_t)\right.$.

In [P1]–[P3] alternative approaches are studied because the EKF can significantly underestimate the posterior covariance in highly nonlinear situations [8] and is unable to capture multiple peaks in the likelihood function. The problems studied in [P2] and [P3] suffer from both; their measurement functions are highly nonlinear and their likelihood functions have multiple peaks.

Another weakness of the EKF is that the computation of the Jacobians can be numerically unstable and of high computational cost [12]. In order to circumvent these potential problems, point-based filters have been developed. Those are derivative-free filters that achieve for nonlinear functions higher order approximations than the EKF [41].

Unscented Kalman filter

One well-known derivative-free filter is the unscented Kalman filter (UKF), which is based on the unscented transformation (UT) [44, 45].

Let $\underline{\mathbf{x}} \sim \mathrm{MVN}(\hat{\mathbf{x}}, \mathbf{P})$, and $\underline{\mathbf{y}} = \mathbf{g}(\underline{\mathbf{x}})$ with nonlinear function $\mathbf{g}$. The idea of UT is to capture the mean and covariance of the distribution of $\underline{\mathbf{x}}$ using a set of deterministically chosen (so-called) sigma points. The mean and covariance of $\underline{\mathbf{y}}$ can then be estimated by propagating the sigma points through $\mathbf{g}$.

For SSM (2) the UKF works as follow. Let $\mathrm{MVN}(\hat{\mathbf{x}}_{t-1}, \mathbf{P}_{t-1})$ be the posterior at $t-1$. In the prediction step the sigma points[1] are formed for $i = 1, \ldots, n_{\mathbf{x}}$ by

$$\begin{aligned} \mathcal{X}_{t-1}^{(0)} &= \hat{\mathbf{x}}_{t-1}, \\ \mathcal{X}_{t-1}^{(i)} &= \hat{\mathbf{x}}_{t-1} + \sqrt{n_{\mathbf{x}} + \gamma}\left[\sqrt{\mathbf{P}_{t-1}}\right]_i, \\ \mathcal{X}_{t-1}^{(i+n_{\mathbf{x}})} &= \hat{\mathbf{x}}_{t-1} - \sqrt{n_{\mathbf{x}} + \gamma}\left[\sqrt{\mathbf{P}_{t-1}}\right]_i, \end{aligned} \tag{8}$$

where $\left[\sqrt{\mathbf{P}_{t-1}}\right]_i$ denotes the ith column of matrix square root of $\mathbf{P}_{t-1}$ and $\gamma = \zeta^2(n_{\mathbf{x}} + \kappa) - n_{\mathbf{x}}$. The parameters ζ and κ determine the spread of the sigma points around $\hat{\mathbf{x}}_{t-1}$ [84]. Then the sigma points are propagated through the state transition function to get

$$\hat{\mathcal{X}}_t^{(i)} = \mathbf{f}_{t-1}\left(\mathcal{X}_{t-1}^{(i)}\right) \tag{9}$$

for $i = 0, \ldots, 2n_{\mathbf{x}}$. The predicted mean $\hat{\mathbf{x}}_t^-$ and covariance $\mathbf{P}_t^-$ are computed as

$$\hat{\mathbf{x}}_t^- = \sum_{i=0}^{2n_{\mathbf{x}}} \omega_i^{(m)} \hat{\mathcal{X}}_t^{(i)}, \tag{10a}$$

$$\mathbf{P}_t^- = \sum_{i=0}^{2n_{\mathbf{x}}} \omega_i^{(c)} \left(\hat{\mathcal{X}}_t^{(i)} - \hat{\mathbf{x}}_t^-\right)\left(\hat{\mathcal{X}}_t^{(i)} - \hat{\mathbf{x}}_t^-\right)^T + \mathbf{Q}_{t-1}, \tag{10b}$$

where the weights $\omega_i^{(m)}$ and $\omega_i^{(c)}$ are defined as [84]

$$\omega_0^{(m)} = \frac{\gamma}{n_{\mathbf{x}} + \gamma}, \quad \omega_0^{(c)} = \frac{\gamma}{n_{\mathbf{x}} + \gamma} + (1 - \zeta^2 + \rho)$$

[1] Note that the presented method for forming sigma points is the most popular. However, there exist alternative ways for forming these points.

$$\omega_i^{(m)} = \omega_i^{(c)} = \frac{1}{2(n_{\mathbf{x}}+\gamma)}, \qquad i = 1,\ldots,2n_{\mathbf{x}}. \tag{11}$$

Parameter ρ enables to incorporate prior information on the state's distribution. For Gaussian distributions $\rho = 2$ is optimal [84].

In the update step the sigma points are formed for $i = 1,\ldots,n_{\mathbf{x}}$ by

$$\begin{aligned}
\mathscr{X}_t^{-(0)} &= \hat{\mathbf{x}}_t^-, \\
\mathscr{X}_t^{-(i)} &= \hat{\mathbf{x}}_t^- + \sqrt{n_{\mathbf{x}}+\gamma}\left[\sqrt{\mathbf{P}_{t-1}^-}\right]_i, \\
\mathscr{X}_t^{-(i+n_{\mathbf{x}})} &= \hat{\mathbf{x}}_t^- - \sqrt{n_{\mathbf{x}}+\gamma}\left[\sqrt{\mathbf{P}_{t-1}^-}\right]_i.
\end{aligned} \tag{12}$$

Then the sigma points are propagated though the measurement function to get

$$\hat{\mathscr{Y}}_t^{(i)} = \mathbf{h}_t\left(\mathscr{X}_{t-1}^{(i)}\right) \tag{13}$$

for $i = 0,\ldots,2n_{\mathbf{x}}$. Now, predicted mean $\boldsymbol{\mu}_t$, predicted covariance of the measurement $\mathbf{S}_t$, cross-covariance of state measurement $\mathbf{C}_t$, and filter gain $\mathbf{K}_t$ are computed by

$$\begin{aligned}
\boldsymbol{\mu}_t &= \sum_{i=0}^{2n_{\mathbf{x}}} \omega_i^{(m)} \hat{\mathscr{Y}}_t^{(i)}, \\
\mathbf{S}_t &= \sum_{i=0}^{2n_{\mathbf{x}}} \omega_i^{(c)} \left(\hat{\mathscr{Y}}_t^{(i)} - \boldsymbol{\mu}_t\right)\left(\hat{\mathscr{Y}}_t^{(i)} - \boldsymbol{\mu}_t\right)^T + \mathbf{R}_t, \\
\mathbf{C}_t &= \sum_{i=0}^{2n_{\mathbf{x}}} \omega_i^{(c)} \left(\mathscr{X}_t^{-(i)} - \hat{\mathbf{x}}_t^-\right)\left(\hat{\mathscr{Y}}_t^{(i)} - \boldsymbol{\mu}_t\right)^T, \\
\mathbf{K}_t &= \mathbf{C}_t \mathbf{S}_t^{-1}.
\end{aligned} \tag{14}$$

Finally, the posterior mean $\hat{\mathbf{x}}_t$ and posterior covariance matrix $\mathbf{P}_t$ can be computed by

$$\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_t^- + \mathbf{K}_t\left(\mathbf{y}_t - \boldsymbol{\mu}_t\right), \tag{15a}$$

$$\mathbf{P}_t = \mathbf{P}_t^- - \mathbf{K}_t \mathbf{S}_t \mathbf{K}_t^T. \tag{15b}$$

One advantage of the UKF is that the number of sigma points grows linearly with respect to the state dimension $n_{\mathbf{x}_t}$ [21]. However, one drawback of the UKF is that it is unstable because for $n_{\mathbf{x}_t} > 3$ the covariance matrix is not always positive definite [11, 12].

Alternative Gaussian filters

The EKF and the UKF belong to the family of Gaussian filters [12, 39]. This group of filters approximates both prediction $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ and update $p(\mathbf{x}_k|\mathbf{y}_{1:t})$ by a Gaussian with same mean and covariance as the both distributions [39].

Means and covariances of the approximative Gaussians are defined by integrals, which can be evaluated using numerical integration. A well-known numerical integration rule is the Gauss-Hermite quadrature rule (see [74, pp. 99 ff.] for explanation of Gauss-Hermite integration). In [39] Ito and Xiong use it inside a filter they call Gauss-Hermite filter (GHF), which is sometimes also called Gauss-Hermite Kalman filter (GHKF) or quadrature Kalman filter (QKF). The GHF works similar to the UKF, but uses a different set of sigma points. It computes the one-dimensional roots of a pth order Hermite polynomial. The multi-dimensional sigma points are then the Cartesian product of the roots [74, p. 103]. Another difference compared with the UKF is that the weights in the GHF have to be positive, while in the UKF negative weights are possible [12]. For SSM (2) the GHF algorithm can be found, for example, in [74, pp. 103 ff.].

The GHF may fail for heavy-tailed or multimodal posterior or likelihood functions [12], a weakness it shares with the EKF. Another weakness of the GHF is that its computational cost increases exponentially with respect to the state dimension [11]. Arasaratnam and Haykin therefore propose to use the third order spherical cubature rule (see [74, pp. 106 ff.] for details) for approximating multi-dimensional integrals inside their cubature Kalman filter (CKF) [11]. The third order spherical cubature approximation's computational cost increases linearly with respect to the state dimension [11]. It is a special case of the UT with parameters $\zeta = 1$, $\rho = 0$ and $\kappa = 0$ [74, p. 108].

In order to ensure that the covariance matrices are always symmetric and positive definite Arasaratnam and Haykin introduce in [11] also the square-root CKF (SCKF). The SCKF propagates square-root factors of prediction and update error covariance matrices instead of the matrices. Hence, matrix square-rooting is avoided, ensuring that the filter is stable [11]. The square-root approach has been proposed

earlier for the QKF in [10], and the resulting filter is consequently called square-root QKF (SQKF). The SQKF is shown to be algebraically equivalent to the QKF. The only difference is that it propagates the square-root of the covariance matrix instead of the covariance matrix itself. In the paper it is pointed out that also the QKF theoretically preserves the symmetry of the covariance matrix but that in practice numerical errors can cause the covariance to become asymmetric. For the UKF the square-root approach is impossible due to its possibly negative sigma point weights [11].

Another point-based algorithm, is the sparse-grid quadrature filter (SGQF), which is based on the sparse-grid quadrature rule [41]. In the algorithm multi-dimensional integrals are approximated as follows. In the first step quadrature points and weights are selected using univariate moment matching (e.g. using Gauss-Hermite quadrature rule). Using the sparse-grid method the quadrature points and weights are extended in the second step to the $n_{\mathbf{x}}$th dimension. The SGQF has the advantage of being more flexible than, for example, the GHF due to tuneable parameters. Furthermore, the number of sparse-grid quadrature points is a polynomial of $n_{\mathbf{x}}$ [41].

Mixture filters

This thesis considers cases in which measurement models are nonlinear and measurement errors are non-Gaussian. Gaussian filters handle nonlinear measurement models but do not handle non-Gaussian errors. Thus, an alternative approaches has to be chosen. One way to handle such cases is the use of GM models.

Definition 1 (Gaussian Mixture) *A Gaussian mixture is a convex combination of N Gaussian density functions* $MVN\left(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right)$*, namely*

$$p(\mathbf{x}) = \sum_{j=1}^{N} \alpha_j MVN_{\boldsymbol{\Sigma}_j}^{\boldsymbol{\mu}_j}(\mathbf{x}), \tag{16}$$

where $\alpha_j \geq 0$ *and* $\sum_{j=1}^{N} \alpha_j = 1$.

The GM $p(\mathbf{x})$ is a valid density function [9, 79]. The idea behind the GM concept is that any density function can be approximated arbi-

trarily closely, except at discontinuities, by a convex combination of Gaussian densities [5, 57, 79]. The concept has been widely applied.

For example, in [39] a filter is proposed in which the posterior $p_{\underline{\mathbf{x}}_{t-1}|\underline{\mathbf{y}}_{1:t-1}}(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$ is approximated by a weighted sum of multivariate Gaussians. Then a Gaussian filter is applied on each of the Gaussian components. In [12] the Gaussian sum-QKF is proposed, which is based on the QKF and models prior and measurement likelihood function as GMs. The use of a GM filter, which is a bank of Gaussian filters, is recommended for highly nonlinear models in [6]. In [4] parallel planes are used to divide the state space and then approximate the posterior in each subspace by one Gaussian.

A similar idea is used in [21] to reduce the computational cost of the QKF. The state space is divided into subspaces and the QKF is then applied to each subspace. The filter is called multiple QKF (MQKF). In the paper it is shown that for independent subspaces MQKF and QKF are equivalent, but otherwise some information is lost by using the MQKF. However, the tests in [21] show only minor performance degradation while the computational cost decreased significantly due to reduced number of used quadrature points. Finally, the authors point out that the multiple filtering approach is not useful for UKF and CKF, because it would not decrease the amount of used points.

In [36] progressive Gaussian filtering is proposed. The idea is to include measurement information continuously into the known prior estimate, although they are only observed at discrete time steps. According to the authors the approach allows tracking "mean and covariance of the true posterior density by its Gaussian approximation", which enables the noise to be of arbitrary structure. For progressive Gaussian filtering the update step has to be modified.

The modified update step is called progressive update step and consists of two substeps. First, a progressive likelihood with an artificial time that takes values between zero and one (this is the time between two time steps in the discrete filtering) is introduced. In [36] the exponentiation of the (discrete) likelihood is used as progressive likelihood, where the time is the exponent. Second, the filter step is continuously executed to compute a modified posterior that

depends also on the artificial time. This modified posterior is Gaussian for artificial time zero. For the remaining artificial time it is approximated by a Gaussian distribution. Thus, for continuous time the posterior is approximated by a sequence of Gaussians. Given the parameters of these Gaussians appropriate parameters for the discrete-time posterior, which is a Dirac mixture approximation, are available.

Non-Gaussian filters

Besides Gaussian mixture filters alternative filters with non-Gaussian error models exist. In [70] a variational Bayes (VB) filter for nonlinear SSMs with t-distributed measurement errors is developed. By introducing a Gamma distributed auxiliary random variable the measurement model can be expressed as Gaussian distribution conditional on the auxiliary variable, which enables the use of Gaussian filtering.

The t-distribution belongs to the family of Gaussian scale mixtures [82]. In [82] a filtering approach based on the sigma point concept for SSMs with measurement errors following a distribution from the Gaussian scale mixture family is proposed. The paper presents for different members of that family distributions that can be mixed with a zero-mean Gaussian random vector to obtain a conditional Gaussian distribution that is equivalent to the original distribution.

Parallel to [P4], [66] has been proposing a filtering algorithm for nonlinear SSMs with skew-t distributed measurement errors. The algorithm uses a VB approximation for the posterior distribution. By using two auxiliary parameters the skew-t measurement model is transformed into a conditional Gaussian distribution.

There exist many more algorithms for Bayesian filtering, such as particle filters. Furthermore, Bayesian smoothing might be of interest for some applications as well. However, since neither particle filtering nor smoothing are of importance for the content of this thesis, the interested reader is referred to [74], which provides an extensive and sound introduction to the field of Bayesian filtering and smoothing.

2.2 Statistical trilateration

The nonlinear measurement equation (2b) is also used in [P4] and [R1]. In those two publications the trilateration problem is analysed under the assumption that the range measurements are time of arrival (TOA) based and the measurement errors are skew-t distributed.

Using statistical trilateration for positioning has been proposed at least as early as 1959. In that year Groginsky showed that the target location could be obtained by trilateration using only range measurements [33]. He assumed the ranges to contain normal distributed errors, and derived a convergent recursion formula for determining the location estimate.

In the statistical formulation of the trilateration problem, used in [P4] and [R1], state $\underline{\mathbf{x}}$ represents the unknown target location. The measurement function is the Euclidean distance between the target and a reference node at a known location $\mathbf{c}_k$. Thus, the $n_\mathbf{y}$ scalar measurements are modelled as

$$\underline{y}_k | \left(\underline{\mathbf{x}} = \mathbf{x}\right) = h_k(\mathbf{x}) + \underline{v}_k = \|\mathbf{x} - \mathbf{c}_k\| + \underline{v}_k \tag{17}$$

for $k \in \{1, \ldots, n_\mathbf{y}\}$. The additive measurement errors $\underline{v}_1, \ldots, \underline{v}_{n_\mathbf{y}}$ are assumed to be mutually independent random variables, and independent of $\underline{\mathbf{x}}$. The PDF of v_k is denoted as $p_{\underline{v}_k}$.

For range measurements that are not based on TOA but, for example, RSS (17) does not necessarily hold. A short overview on lateration techniques and how they compare with angulation techniques, which use angles for positioning, can be found, for example, in [55]. According to the paper range measurements can be derived from TOA, time difference of arrival (TDOA), or RSS measurements, or from the roundtrip time of flight (RTOF) or received signal phase method. The RTOF method measures the time it takes for the signal to travel from the transmitter to the receiver and back; the received signal phase method estimates ranges using carrier phase or phase difference (see [25, 55] for more details).

Given the measurement vector $\mathbf{y} = y_{1:n_\mathbf{y}}$ the posterior PDF of $\underline{\mathbf{x}}$ is

$$p_{\underline{\mathbf{x}}|\underline{\mathbf{y}}}(\mathbf{x}|\mathbf{y}) \propto p_{\underline{\mathbf{x}}}(\mathbf{x}) \prod_{k=1}^{n_\mathbf{y}} p_{\underline{v}_k}\left(y_k - \|\mathbf{x} - \mathbf{c}_k\|\right). \tag{18}$$

In [P4] and [R1] the posterior mode, i.e. the value of $\underline{\mathbf{x}}$ that maximises (18), is determined. The posterior mode is generally called the maximum *a posteriori* (MAP) estimate. For a uniform prior this MAP estimate coincides with the maximum likelihood (ML) estimate.

The algorithms used in [P4] and [R1] for finding MAP estimates use the Jacobian of $\mathbf{h} = h_{1:n_\mathbf{y}}$, which is the $n_\mathbf{y} \times n_\mathbf{x}$ matrix $\mathbf{H}$ whose kth row is the transpose of a unit vector pointing from $\mathbf{c}_k$ to $\mathbf{x}$, that is,

$$\mathbf{H}_{k,1:n_\mathbf{x}} = \frac{\partial h_k(\mathbf{x})}{\partial \mathbf{x}} = \frac{(\mathbf{x}-\mathbf{c}_k)^T}{||\mathbf{x}-\mathbf{c}_k||}. \tag{19}$$

3 Offline phase: Radio map generation

Almost all positioning methods require some sort of additional information about network infrastructure or measurement behaviour. For example, for positioning in a cellular phone network the locations of base stations from which the user equipment receives signals have to be known in order to enable positioning the UE. Those BS locations are not always known, and have to be estimated from measurement data, which is done in the so-called offline phase. In the offline phase a so-called radio map (RM) is generated, which is then used to compute the UE position either on a server or on the UE itself.

This section describes the RMs required for the positioning algorithms presented in Section 5 and how they can be generated. [P1] covers the forms of RMs presented in Subsections 3.1 – 3.3 in more detail. Furthermore, Subsection 3.4 contains a short review on existing alternative RM approaches that are not handled in the thesis' publications.

3.1 Radio map for nonparametric methods

When determining the position of a UE inside a WLAN most algorithms exploit the correlation between the UE's location and the RSS values [61, p. 47]. Since WLAN-based positioning is often used for indoor environments, where modelling the signal propagation is rather complex, so-called nonparametric location fingerprinting methods are widely applied for positioning [38].

The idea behind these methods is to store fingerprint (FP) data in the RM. A FP is a measurement of radio characteristic records taken in a known location. For example, for a WLAN a FP consists of location coordinates, a list of access point identifiers (AP-IDs) for APs from which signals are received in that location, and RSS values of those signals.

In each location the channel conditions can only be assumed constant for a short time [53]. Therefore, Han et al. recommend in [35] to collect 100 samples for a single location in a WLAN (the number of samples typically depends on the used signal type, methods, etc.).

Furthermore, they point out that for indoor WLANs the typical FP density is one FP per square meter. Thus, a large amount of measurements has to be collected and stored in the RM, which is one of the disadvantages of nonparametric methods. For indoor environments those fingerprints generally have to be collected by site survey, meaning that locations have to be entered by hand, which makes it a tedious job. In outdoor environments FP data collection is easier, because the data can be collected via war-driving or crowd-sourcing, which both use GNSS-based location estimates.

The size of the radio map can become another problem for nonparametric methods, because it depends on the number of FPs. To achieve high positioning accuracy in the online phase, i.e. the positioning phase, a high FP density is required, which results in a large RM. So-called parametric methods circumvent this problem by summarising FP data and storing only some parameters in the RM, reducing its size significantly. Possible reductions are presented in [P1] and [62]. In the following two subsections some parametric methods are explained, how they summarise FP data and what they store in their RM.

3.2 Coverage area models

A computationally light parametric method is the coverage area (CA) approach introduced in [49, 50]. The algorithm's idea is to fit an elliptical probability distribution to FP locations in which signals from one specific transmitter are received.

Instead of FPs, only the distribution's parameters are stored in the RM. In [49, 50] the authors choose a multivariate normal distribution for describing the coverage area of WLAN APs. Thus, for two-dimensional FP locations only 5 parameters per AP have to be stored in the RM; 2 for the mean and 3 for the covariance of the bivariate normal distribution. Therefore, the RM size depends on the number of APs, a feature that it shares with other parametric FP methods.

Given FP locations $\mathbf{z} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n\}$ in which the transmitter's signal are received, the CA is modeled in [49, 50] by a posterior distribution

for the ellipse parameters $\underline{\boldsymbol{\theta}}$. This distribution is given by Bayes' rule

$$p_{\underline{\boldsymbol{\theta}}|\underline{\mathbf{z}}}(\boldsymbol{\theta}|\mathbf{z}) \propto p_{\underline{\mathbf{z}}|\underline{\boldsymbol{\theta}}}(\mathbf{z}|\boldsymbol{\theta}) p_{\underline{\boldsymbol{\theta}}}(\boldsymbol{\theta}), \tag{20}$$

where the likelihood and prior are normal. The regression problem's Bayesian formulation allows exploiting information about "typical" sizes of CAs via the prior $p_{\underline{\boldsymbol{\theta}}}(\boldsymbol{\theta})$, which is available through experimental studies or can be computed using equations such as Friis' transmission equation or the d^{-4} power law (see [61, p. 47 ff.] for details). Furthermore, (20) enables updating the posterior as new FP data becomes available [48, p. 14 ff. and p. 29 ff.]

The CA approach is used in [P2] for computing the prior and as a comparative method, and in [P1].

3.3 Path loss models

Although modelling the signal propagation can be challenging, especially for indoor environments, it is widely used for positioning. Generally, so-called path loss (PL) models are used that describe either the signal power loss L_P or the received signal strength P_{RSS} along a radio link, averaged over large-scale and small-scale fading [61, p. 27].

For several real-world applications, assuming the PL model's parameters to be known is an oversimplification and ill-suited [53]. Therefore, methods have been developed to estimate those parameters and transmitter locations from FP data consisting of FP locations, observed transmitter IDs and corresponding RSS values.

Some of these methods estimate first the transmitter location or assume it to be known before estimating the parameters (e.g. [17, 72]); others estimate the transmitter position and parameters simultaneously (e.g. [53, 65]). For the first group using the CA approach from the previous section would be an option for estimating the locations. Using fixed PL model parameters for estimating the transmitter location is not recommended. It has been shown that this approach can cause large errors if the parameter values are chosen inaccurately [53]. Li therefore recommends a joint estimation of location and parameters in [53], and stresses that it removes the necessity of extensive channel measurements.

In [53] the classic narrowband radio propagation PL model is considered. This model has only one parameter, namely the PL exponent, which is also called the distance-power gradient. However, in publication [P2] a version of the log-distance model, namely

$$\underline{P}_{\mathrm{RSS}}(d) = A - 10\eta \log_{10}(d) + \underline{w}, \tag{21}$$

is used. Variable d denotes the distance between transmitter and receiver. In addition to PL exponent η, this model contains the apparent transmission power parameter $A = P_{\mathrm{RSS}}(1)$ that has to be estimated. Shadow fading is modelled by the zero-mean normal random variable $\underline{w}$ with variance σ_w^2.

The two model parameters and transmitter location are estimated in [65] by the Iterative Reweighted Least Squares (IRLS) method, which is a Gauss-Newton approach. Similar to the CA approach, the Bayesian method used in [65] enables updating the posterior as new FP data becomes available and using appropriate priors. According to [65] most care has to be taken for the transmitter location prior; given enough FPs the influence of the PL model parameter prior is negligible.

As for the CA approach, the size of the RM depends on the number of transmitters rather than the number of FPs. In addition, the size is influenced by the used PL model (1 parameter in the model used by Li [53], 2 parameters in the model used by Nurminen et al. [65]), and whether only point estimates or also uncertainty measures for the estimates are stored. For example, the IRLS used in [65] yields a multivariate normal distribution for transmitter location and a bivariate normal distribution for parameters A and η. The covariance matrices of those two normals show how certain the estimates are and can be stored in the RM. Using two normals rather than one leaves the cross-correlation between transmitter location and parameters unknown but simplifies the estimation process. Another trick used in [65] is the adding of a small constant to the diagonal terms of the transmitter location's covariance matrix to account for correlation in FP data.

The PL model approach is used for the GGM in [P2] (denoted GMA in [P2]) but it could also be used in combination with the method

developed and analysed in [P4] and [R1] if the measurements are RSS values rather than ranges.

Besides the log-distance PL model (21) there exist various other PL models that could be used instead. In [35] an extension of Seidel's model [76] is proposed that takes into account the angles between signal path and obstacles, such as walls. A two-slope PL model is used in [19], which uses two different PL models for regions near to and far from the transmitter. The authors choose this approach because for large distances between CN and UE the PL drops faster than for short distances. In [77] the authors go further, and use a multi-slope PL model, whose PL exponent is distance dependent. Furthermore, they extend the model with an additive floor loss parameter. A more detailed discussion can be found in [P1].

Depending on the chosen PL model the RM entries vary significantly. The more sophisticated the model the more parameters need to be stored in the RM. In order to avoid storing PL model parameters in the RM the parameters could be included in the state $\mathbf{x}_t$ and be updated inside the filter that is used for positioning (see [18] for an example). However, in this approach suitable initial values for the PL parameters will be beneficial. Thus, having a RM with values for the parameters might be useful even in this approach.

3.4 Alternative radio map approaches

It should be noted that there exist alternative RMs that are used by different positioning methods, which are not covered in the publications of this thesis.

For example, in recent years the use of geomagnetic fields has been proposed for indoor positioning [25]. In [20] the authors collect information about the magnetic field in the FP location and store them in the RM. Their positioning approach uses their finding that steel and concrete skeletons distort geomagnetic fields inside buildings, which results in significant errors in compass readings that are location dependent but stable in time [20, 52]. The average deviation from the ground truth direction is found to be 45 degree in the tests in [20].

In [52] the authors stress that magnetic interferences have to be con-

Table 1: Comparison of radio maps. Table contains information on references in which the methods have been discussed, what the radio maps contain, and the pros and cons of the different methods.

Method	RM entries	Pros	Cons
Nonparametric methods [38]	FPs (location, transmitter IDs, RSS values	no need to model signal propagation	RM size depends on number of FPs
Coverage area [49, 50]	one Gaussian per transmitter	RM size depends on number of transmitters; CA can be updated using Bayes' rule	probability mass located around centre of CA (can be problematic for weak signals)
Path loss model [65]	two Gaussian per transmitter (one for location & for PL model parameters)	RM size depends on number of transmitters; takes RSS values into account	modelling of signal propagation can be challenging (NLOS, multipath etc.)
Geomagnetic fields [20]	FPs (contain information on magnetic field in FP location)	no multipath & NLOS errors [52]; magnetic field location dependent	RM size depends on number of FPs; magnetic field not unique
Markers (2-dim barcodes) [63]	locations & IDs of markers	RM size depends on number of markers; unique marker IDs	high precision requires high marker density

sidered when using magnetic fields. The authors also point out that a location's magnetic field, although location dependent, is not necessarily unique. This is because for the magnetic field only three components are available if the magnetic north is known (otherwise only two components are available) [25, 52]. In order to remove this ambiguity, in [30] it is proposed to store sequences of magnetic FPs, measured along a known path, in the RM instead of single FPs.

Another alternative is image-based positioning. For example, in [63] a software is introduced that allows to determine the position and orientation of a mobile phone. The method uses a RM containing information on markers, which are two-dimensional barcodes that provide unique IDs, and their locations. Positioning is done by continuously scanning the environment, using the mobile phone's camera,

for the markers.

Table 1 contains a summary of the radio map approaches discussed in this section. For each method a reference is given in which the method has been discussed, and what is stored in the method's RM. In addition, the pros and cons of each method are summarised.

The different RMs discussed in this thesis, except for the RMs for nonparametric methods in Subsection 3.1 and for geomagnetic field-based positioning, all contain information on the location of reference points. A reference point can be a signal transmitter (for CA and PL model approach), a marker (used in [63]), a light source (used in [56]), or a loudspeaker (used in [46]). The locations of reference points can be known exactly or have to be estimated from measurement data.

4 Offline phase: Fitting normal and skew-t distributions to data

In [P3] and [R1] UWB and LTE networks are used to analyse the positioning algorithms developed in [P2] and [P4] respectively. The measurements in those networks are time delay-based range measurements, which makes the use of PL models obsolete. However, because those measurements contain errors, distributions are fitted to ranging error data in [P3] and [R1].

For the UWB network in [P3] for simplicity zero-mean normal distributions are fitted to both line of sight (LOS) and NLOS errors. This way the GGMs for LOS and NLOS are the same except for the value of one parameter. It is well known that NLOS ranging measurements contain a positive bias (aka NLOS error) because the signal has to travel an extra distance to arrive at the receiver [32]. This means, that the zero-mean assumption should not hold for NLOS. However, the equipment used in [P3] did not label ranging errors always correctly as either LOS or NLOS, and most NLOS ranging errors were concentrated around zero (see lower plot in Figure 2). Hence, it was decided to neglect the NLOS error in [P3].

In [P3] it is shown that for highly accurate ranges the GGM, which assumes zero-mean normal distributed ranging errors, is ill-suited. Therefore, in [R1] normal distributions that are not necessarily zero-mean and skew-t distributions are fitted to ranging errors from different channels of the LTE network. In all cases the fitted skew-t distributions resemble the empirical error distribution more closely than the fitted normal distributions. The parameters of normal distributions are fitted by maximum likelihood. Fitting the parameters of a skew-t distribution $\mathrm{ST}\left(\xi, \sigma^2, \lambda, \nu\right)$ is more complicated, and is done in [P4] and [R1] by a Gibbs sampler [31].

The idea of sampling is to simulate samples from a distribution of interest, which are then used to approximate integrals (e.g. the integral that defines the mean value of ξ). If the number of samples converges to infinity then sampling methods yield approximation results that are equivalent to the results of integration [58, pp. 79 ff.]. For example, the mean of parameter ξ can be obtained by either evaluating the integral defining the mean value or by generating a large

amount of samples for ξ and use their average as mean estimate.

Let $\mathbf{v} = v_{1:n}$ be a vector of n independent ranging errors that are assumed to be skew-t distributed, i.e. $\underline{v}_j \sim \mathrm{ST}\left(\xi, \sigma^2, \lambda, \nu\right)$. Sampling from the joint posterior $p\left(\xi, \sigma^2, \lambda, \nu | \mathbf{v}\right)$ is unfeasible. Instead the idea of the Gibbs sampler (GS) is to sample from the conditional posterior distributions of the parameters. This has the advantage that it suffices if the conditional density of a parameter is known up to a normalising constant [58, p. 99]. In [29] it is shown that by using an alternative parametrisation the parameters can be estimated using MCMC samplers, such as the GS. The conditional posterior distributions from [29] can be found in the paper's supplement document.

Input: μ_ξ, σ_ξ, s_r, θ_r, μ_λ, σ_λ, $\nu_{\max}$, n (no. of ranging errors in data)
Draw ξ from $\xi \sim \mathrm{N}(\mu_\xi, \sigma_\xi^2)$
Compute $\sigma^2 \leftarrow 1/r$ with $r \sim \mathrm{Gamma}(s_r, \theta_r)$
Draw λ from $\lambda \sim \mathrm{N}(\mu_\lambda, \sigma_\lambda^2)$
Compute $\delta_\lambda \leftarrow \lambda/\sqrt{1+\lambda^2}$
Draw ν from $\nu \sim \mathrm{U}(2, \nu_{\max})$
for $j = 1$ *to* n **do**
    Draw τ_j from $\tau_j \sim \mathrm{Gamma}(\nu/2, \nu/2)$
    Compute $t_j \leftarrow |\bar{t}|$ with $\bar{t} \sim \mathrm{N}(0, \sigma^2/\tau_j)$
    Draw y_j from $y_j \sim \mathrm{N}(\xi + \delta_\lambda t_j, \frac{1-\delta_\lambda^2}{\tau_j}\sigma^2)$
end
Draw τ_{pred} from $\tau_{\mathrm{pred}} \sim \mathrm{Gamma}(\nu/2, \nu/2)$ `// predicted` τ
Compute $t_{\mathrm{pred}} \leftarrow |\bar{t}|$ with $\bar{t} \sim \mathrm{N}(0, \sigma^2/\tau_{\mathrm{pred}})$ `// predicted` t
Draw y_{pred} from $y_{\mathrm{pred}} \sim \mathrm{N}\left(\xi + \delta\, t_{\mathrm{pred}}, \frac{1-\delta_\lambda^2}{\tau_{\mathrm{pred}}}\sigma^2\right)$ `// predicted observation`

Algorithm 1: Model used by Gibbs sampler for fitting skew-t

The GS works as follows. After assigning initial values to the skew-t distribution's 4 parameters, they are ordered and updated by drawing samples from their conditional distributions given $\mathbf{v}$ and current estimates of the 3 other parameters. This process is repeated $T_0 + T$ times. The first T_0 samples for each parameter are called "burn-in" samples and are discarded; the sample means of the last T samples are used as parameter estimates. The sample standard deviation can be used to evaluate how certain the parameter estimates are.

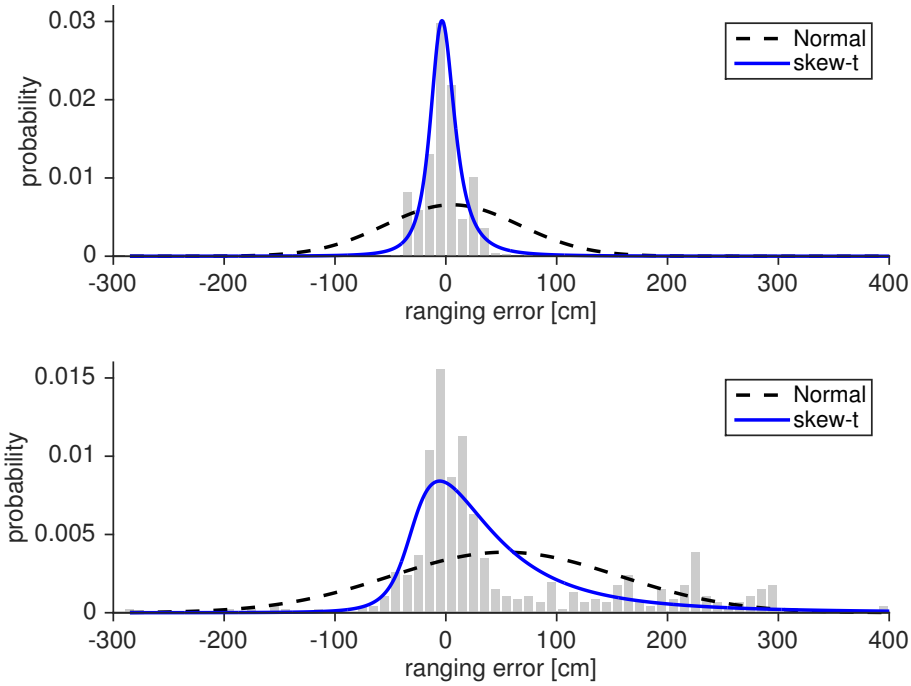


Figure 2: Histograms of LOS (upper plot) and NLOS ranging error data from [P3] with fitted normal and skew-t distributions. The sample mean and the square root of the unbiased estimator of the variance are used as mean and standard deviation for the fitted normal distributions; the parameters of the skew-t distributions are estimated using a Gibbs sampler with $T_0 = 1\,000$ and $T = 10\,000$.

Gibbs sampling has the advantage that software can be designed that works for almost any model [64, p. 846]. For estimating the parameters the hierarchical model defined by (58) is used for the jth ranging error. The estimations in [P4] and [R1] use JAGS [1], which is a program that allows analysing Bayesian hierarchical models by Gibbs samplers. Algorithm 1 shows the model used in the GS implementation for fitting parameters of skew-t distribution. For all 4 parameters diffuse priors are used; ξ and λ have normal priors, $\sigma^2 = 1/r$ with r having a Gamma-distributed prior, and ν has a uniform prior with lower limit greater than 2 to ensure that mean and variance of the skew-t distribution exist. Given these parameter estimates Algorithm 1 then generates predicted observations, which

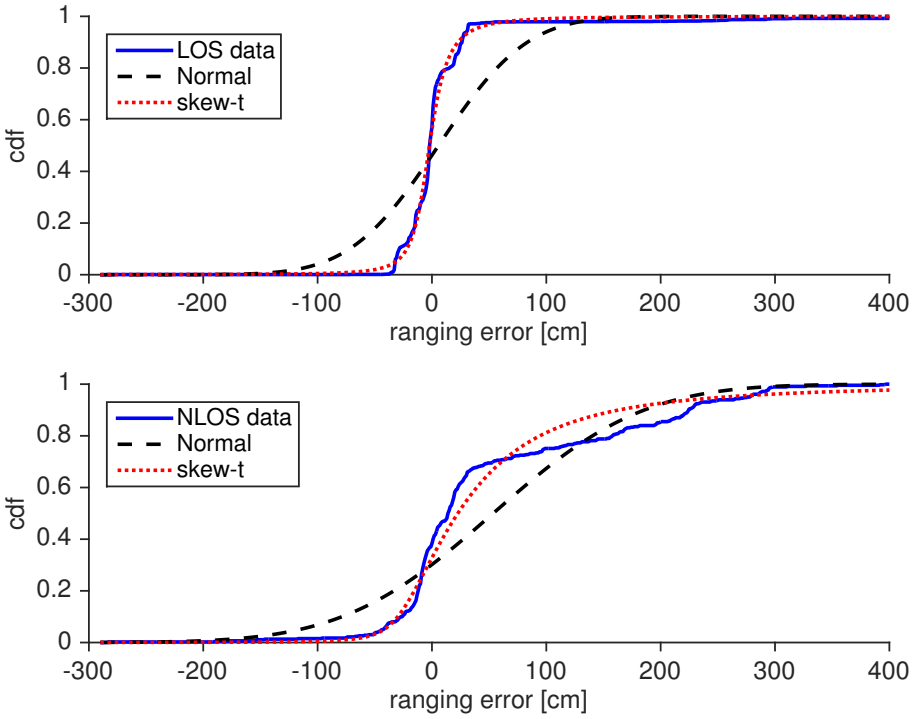


Figure 3: Empirical CDFs of LOS (upper plot) and NLOS ranging error data from [P3] with CDFs of fitted normal and skew-t distributions. The sample mean and the square root of the unbiased estimator of the variance are used as mean and standard deviation for the fitted normal distributions; the parameters of the skew-t distributions are estimated using a Gibbs sampler with $T_0 = 1\,000$ and $T = 10\,000$.

can be compared with $\mathbf{v}$ to evaluate the quality of fit with current estimates. The model in Algorithm 1 has to be only passed to JAGS, which then generates the samples [64, p. 847]. The implementation of the GS has been published for [P4].[2]

In [P4] it is shown by simulations that for a set of 1 000 measurements $T_0 = 200$ and $T = 500$, suffices to provide reliable parameter estimates. Increasing the number of samples to $T_0 = 1\,000$ and $T = 10\,000$ does neither improve the estimation accuracy nor reduce the uncertainty significantly.

[2] The Gibbs sampler implementation is available for download at https://PMullerTUT@bitbucket.org/PMullerTUT/trilaterationskewterrors.git

Figure 2 shows the histograms of LOS and NLOS UWB data from [P3] as well as fitted normal and skew-t distributions. The distributions' parameters have been estimated using the GS from [P4]. Figure 3 shows the empirical cumulative distribution functions (CDFs) for the LOS and NLOS UWB data as well as the CDF's of the fitted distributions. For both data sets the fitted skew-t resamples the UWB data, i.e. the histograms an the empirical CDFs, much closer than the fitted normal.

5 Online phase: positioning

Once a RM is generated it is possible to apply a variety of algorithms for positioning the UE. This section focuses on the methods used in [P2]–[P4] and [R1], and summarises the findings of those publications. Alternative approaches are only briefly discussed.

According to [55] positioning techniques can be divided into three groups: triangulation, scene analysis and proximity algorithms. The GGM, the dGN, the EM and the GS algorithms proposed and analysed in [P1]– [P4] and [R1] belong to the first group, more precisely to its subgroup of lateration techniques. Lateration infers the position by measuring distances to reference points or computing them using RSS or time-delay measurements [55]. Angulation techniques, which are the second subgroup of triangulation techniques, use angles relative to reference points rather than distances for positioning. These techniques have the advantage that two or three measuring devices are sufficient for two- or three-dimensional positioning respectively. However, the hardware requirements are relatively large and complex. Furthermore, the accuracy degrades as the distance between UE and reference points increases [55].

The methods discussed in Subsection 5.1 belong to the group of scene analysis techniques. These methods collect features (aka fingerprints) in known locations in the offline phase and then infer the UE position in the online phase by matching the measurement from the unknown location with the RM containing the FPs [55]. The group contains, for example, nearest neighbour and probabilistic methods (both of which are discussed in Subsection 5.1), neural networks, support vector machines and smallest M-vertex polygon method. More details and references where these techniques have been applied for positioning can be found in [55].

Proximity algorithms are generally simple to implement but provide only symbolic relative position estimates. The idea is to consider the UE to be collocated with the transmitter from which it receives a signal or by which it is detected [25, 55]. The coverage area approach in Subsection 5.2 can be interpreted as a (sophisticated) proximity technique, because it only uses information whether the UE receives a signal from a certain transmitter but ignores the RSS.

5.1 Positioning with nonparametric methods

The radio map for nonparametric methods consists of FP data (see Subsection 3.1 for details). In the positioning phase the UE's measurements are compared with the RM's FP entries to infer a position estimate [38]. There exist various approaches for determining this estimate, which can be divided in deterministic and probabilistic methods. Here only a brief overview on theses methods is given. For additional information and references the reader is referred to [38], on which this Subsection is based.

Deterministic nonparametric methods

Deterministic nonparametric positioning methods assume the state to be deterministic (hence their name). The estimate $\hat{\mathbf{x}}$ is a convex combination of N FP locations $\mathbf{l}_{\mathrm{FP}_j}$, this means

$$\hat{\mathbf{x}} = \sum_{j=1}^{N} \frac{\omega_j}{\sum_{j=1}^{N} \omega_j} \mathbf{l}_{\mathrm{FP}_j}, \tag{22}$$

where $\omega_j \geq 0$ for all j. Let $\mathbf{y}$ be the vector containing the measured RSS values and $\mathbf{P}_{\mathrm{RSS}_j}$ be the vector containing the RM-stored RSS values of each transmitter in the jth FP location. A possible weight ω_j is the inverse of the norm of RSS innovation that is defined as

$$\omega_j = \frac{1}{||\mathbf{y} - \mathbf{P}_{\mathrm{RSS}_j}||}. \tag{23}$$

For the norm $||\cdot||$, for example, Manhattan norm (1-norm), Euclidean norm (2-norm), and Mahalanobis norm are used (see [38] for definitions of various norms and references in which they are used). In [77] it is pointed out that using (23) finding $\hat{\mathbf{x}}$ can be interpreted as a minimisation problem, because a FP location's weight is highest if the sum of differences between observed transmitter RSS values in current UE location and that FP location is minimal. An alternative weight could be the number of transmitters that are observed in both current UE and FP locations. This approach can be interpreted as a maximisation problem.

The simplest approach based on (22), the nearest neighbour (NN) method, uses the location of the FP with the highest weight as position estimate. In [P1] the weighted K-nearest neighbour (WKNN)

algorithm, an extension of NN, is used for comparison. The WKNN computes the UE position estimate according to (22) using only the FP locations with the K largest weights. For the remaining FP locations the weights are set to zero. The WKNN is widely applied because it yields good robustness and accuracy while having only medium complexity and cost [55].

The WKNN performs at least as well as all parametric methods for a test in real-world indoor WLANs in [P1] when all data is used. In the test area, two buildings at Tampere University of Technology, the WLAN AP density is high. However, when removing 90% of the APs from the test area the WKNN positioning accuracy deteriorates significantly and the algorithm is outperformed by the parametric methods. This is in line with [28], where it was discovered that range-based positioning techniques outperformed nonparametric FP positioning algorithms in case of limited FP data.

Probabilistic nonparametric methods

Probabilistic nonparametric methods assume the state $\underline{\mathbf{x}}$ to be random, and the position estimate is computed using Bayes' rule (1). The idea of probabilistic methods is to divide the area covered by the RM into cells. A natural choice is to use the FP locations as centres of those cells, resulting in N cells $b_1, \ldots, b_N$.

In [38] a uniform prior is used. Hence the prior is

$$p(\mathbf{x}) = \frac{\sum_{j=1}^{N} [\mathbf{x} \in b_j]}{\sum_{j=1}^{N} |b_j|,} \tag{24}$$

where $[\cdot]$ is the Iverson bracket and $|b_j|$ is the volume of the jth cell. Furthermore, the measurements from one FP are assumed to represent the RSS distribution inside the whole cell. Hence the likelihood inside that cell is constant, and the likelihood is

$$p(\mathbf{y}|\mathbf{x}) = \sum_{j=1}^{N} p(\mathbf{y}|b_j)[\mathbf{x} \in b_j]. \tag{25}$$

The likelihood $p(\mathbf{y}|b_j)$ can be computed by several approaches. In [38] suitable methods and references in which they are used are given.

One obvious drawback of probabilistic nonparametric methods is the need to compute the likelihood in each of the N FP locations stored in the RM, which can be computationally demanding. Therefore, in the following subsections (probabilistic) parametric methods are considered.

Another issue that affects all FP positioning methods discussed in this subsection is the generation and maintenance of the RM. When generating the RM data might not be collected in some areas, because they are restricted or inaccessible, which results in gaps in the RM [80]. Furthermore, one of the tests with real-world WLAN data in [P1] shows that FP-based positioning methods (represented by the WKNN in [P1]) suffer significant performance deterioration when the RM is outdated, because the radio environment is constantly changing [38, 86].

In [80] various interpolation and extrapolation techniques for recovering missing FP data and filling gaps in the RM are studied using extensive WLAN data. For recovering missing data techniques from two categories are considered. In the first category, linear interpolation is used to fill the gaps between known FP locations while for extrapolation minimum method, mean method and gradient method are used to estimate the data in the gaps. In the second category, interpolation and extrapolation are carried out jointly using NN method and inverse distance weighting.

The methods studied in [80] are mainly meant for RM generation. However, they could also applied for RM maintenance. For example, instead of collecting new FPs for all FP locations in the RM new data could be collected only in some of them and for the remaining locations FP data could be updated using the presented interpolation and extrapolation techniques.

5.2 Coverage area models

The CA-based positioning method is used in [P2], for comparison, and in [P1]. It is developed and explained in detail for positioning in [48]. The method requires a RM with estimated CAs of transmitters. In [P1] and in [P2] the CA of a transmitter is modelled as a multivariate normal distribution with place $\boldsymbol{\mu}$ and shape $\boldsymbol{\Sigma}$. The CA-based posi-

tioning method uses only a list $\mathbf{ID} = \{\mathrm{ID}_1, \mathrm{ID}_2, \ldots, \mathrm{ID}_K\}$ of transmitter IDs observed by the UE in its current location $\mathbf{x}$.

Under the assumption of mutually independent observations the likelihood is

$$\begin{aligned} p(\mathbf{ID}|\mathbf{x}) &= \prod_{k=1}^{K} p(\mathrm{ID}_k \in \mathbf{ID}|\mathbf{x}) \\ &\propto \exp\left(-\frac{1}{2}\sum_{k=1}^{K}(\mathbf{x}-\boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}-\boldsymbol{\mu}_k)\right) \\ &= \exp\left(-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^T \mathbf{S}^{-1}(\mathbf{x}-\bar{\mathbf{x}}) + \text{constant}\right) \end{aligned} \tag{26}$$

with $\mathbf{S} = \left(\sum_{k=1}^{K} \boldsymbol{\Sigma}_k^{-1}\right)^{-1}$ and $\bar{\mathbf{x}} = \mathbf{S}\left(\sum_{k=1}^{K} \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k\right)$.

Koski [48, p. 35] points out that the assumption of mutual independence is a weakness of the proposed approach and might not always hold. For example, CAs of neighbouring transmitters usually overlap. Therefore, assuming that observing ID_i is independent of observing its neighbour ID_j does not hold. However, the assumptions was chosen in order to keep the algorithm simple and computationally light.

The position estimate of the CA-based method given the prior $\underline{\mathbf{x}} \sim \mathrm{MVN}\left(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0\right)$ is

$$\begin{aligned} p(\mathbf{x}|\mathbf{ID}) &\propto p(\mathbf{x})p(\mathbf{ID}|\mathbf{x}) \\ &\propto \exp\left(-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}}_0)^T \mathbf{S}_0^{-1}(\mathbf{x}-\bar{\mathbf{x}}_0)\right), \end{aligned} \tag{27}$$

which is a multivariate normal with covariance $\mathbf{S}_0 = \left(\sum_{k=0}^{K} \boldsymbol{\Sigma}_k^{-1}\right)^{-1}$ and mean $\bar{\mathbf{x}}_0 = \mathbf{S}_0\left(\sum_{k=0}^{K} \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k\right)$. The prior can be interpreted as one CA itself, because it is also a multivariate normal like the CAs themselves. Then the position estimate is the weighted mean of the CA centres with weights being the inverses of their covariance matrices. The posterior covariance matrix provides an uncertainty measure for the estimate; a measure that is not provided by the WKNN algorithm.

In case no prior information on the location is available, Koski [48, p. 35] suggest to use a prior with large covariance. Furthermore, she

describes how information that a transmitter is not observed in the UE location could be used for positioning but points out that this will become impractical (due to high computational demand) for large radio maps.

In the WLAN-based tests of [P1] the CA approach performs well in all 4 test scenarios. When using all data and up-to-date RM its accuracy is only slightly worse than those of WKNN and the more complicated parametric methods. For the scenario where only APs with 5 strongest RSS values are used for positioning and in the scenario with low AP density it still performs on a similar level and close to the other parametric methods. Finally, it is amongst the best methods for the test scenario with outdated RM; its performance is on the same level as with up-to-date RM.

Thus, while being a quite simple method it provides reasonable precise position estimates under all tested circumstances, making it a good choice as reference algorithm in [P2].

Both [P1] and [P2] also use a filtered version of the CA approach. The update is done using a standard Kalman filter.

5.3 Generalised Gaussian mixture

Modelling a transmitter's coverage area as an ellipse via a multivariate normal has the disadvantage that most of the probabilty mass is located in close proximity of the ellipse's centre. A low RSS, however, indicates in general that the UE is close to the CA's edge. Hence, in cases with significant amount of weak received signals the CA approach's positioning performance is poor [50].

In order to address this issue the measured RSS values should be taken into account. One approach is to use PL models, which have been described in Subsection 3, to compute estimates for the distances between transmitter and UE from the RSS values. For example, let y_k be the kth RSS measurement, where k is not necessarily the index of the transmitter as in the previous subsection. Then using PL model (21) RSS measurement y_k is

$$y_k = A - 10\eta \log_{10}(d_k) + \underline{v} = A - 10\eta \log_{10}(||\mathbf{x} - \mathbf{c}_k||) + \underline{v}, \qquad (28)$$

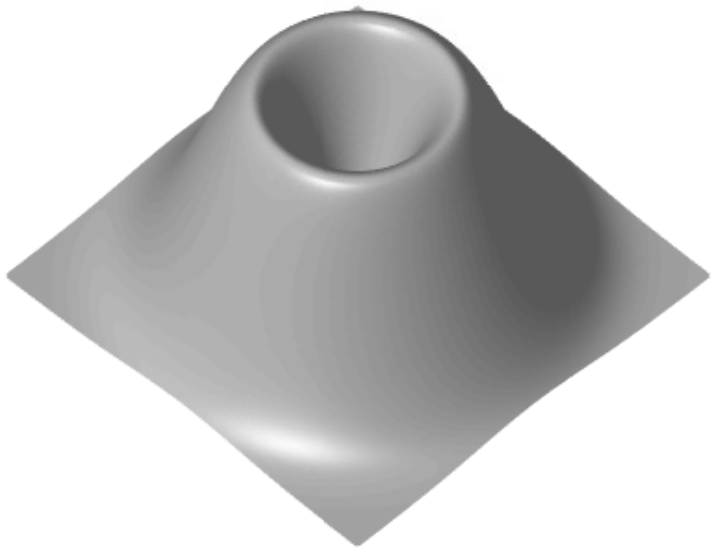


Figure 4: Exact normalised likelihood for a two-dimensional isotropic ranging model. The transmitter is in the centre of the volcano-shaped likelihood and the "volcano's" brow corresponds to a circle around the transmitter location using the range measurement as radius.

where $||\mathbf{x}-\mathbf{c}_k|| = \sqrt{\sum_{i=1}^{n_\mathbf{x}} \left(x_i - c_{k,i}\right)^2}$ is the Euclidean distance d_k between UE location $\mathbf{x}$ and transmitter location $\mathbf{c}_k$.

Given the distance estimates $\mathbf{d} = \{d_1, d_2, \ldots, d_K\}$ the trilateration problem (18) can be solved using various methods. For example, in [65] a grid method that uses standard Monte Carlo integration, the Metropolis-Hastings algorithm, and the IRLS are used. A more thorough discussion of these methods can be found in [P1].

In [P3] instead of RSS values range measurements based on two-way time-of-flight (TW-TOF) are used, thus the measurement model for one range measurement is

$$y_k = ||\mathbf{x}-\mathbf{c}_k|| + \underline{v}, \tag{29}$$

with ranging error $\underline{v}$.

Both (28) and (29) are nonlinear, and using the Kalman filter is not possible; instead the EKF can be used. However, in highly nonlinear

conditions the EKF can seriously underestimate the posterior covariance [8]. Furthermore, the measurement models (28) and (29) are multimodal. Figure 4 shows the exact normalised measurement likelihood $p(y_k|\mathbf{x})$ for a two-dimensional isotropic ranging model, which was used in [P3]. The transmitter is in the centre of the volcano-shaped likelihood, and the "volcano's" brow corresponds to a circle around the transmitter location with range measurement being used as radius and can be interpreted as a infinite number of peaks. The EKF would follow for such multimodal systems only a single peak of the likelihood and therefore yields a maximum likelihood estimate rather than a minimum variance estimator [9].

One way to capture multiple peaks of the likelihood in Figure 4 is to approximate it by a GM, namely

$$p(y_k|\mathbf{x}) \approx \sum_{j=1}^{N} \alpha_j \mathrm{MVN}_{\boldsymbol{\Sigma}_j}^{\boldsymbol{\mu}_j}(m_j(y_k)), \tag{30}$$

where $\alpha_j \geq 0$ and $\sum_{j=1}^{N} \alpha_j = 1$, and $\{m_j\}_{j=1}^{N}$ are some known functions. Figure 5 shows the exact measurement likelihood $p(y_k|\mathbf{x})$ and a GM approximation that uses $N = 5$ Gaussians. The idea of the GM approximation is to pick N peaks of the exact likelihood and use them as mean values $\{\boldsymbol{\mu}_j\}_{j=1}^{N}$. Then the covariance matrices $\{\boldsymbol{\Sigma}_j\}_{j=1}^{N}$ are adjusted such that the resulting GM resembles closely the exact likelihood. However, due to the infinite number of peaks in the exact likelihood a large number of components is required (the 5 Gaussians used in Figure 5 give only a rough approximation of the likelihood). This results in excessive computational complexity [9, 79]. Many component reduction methods have therefore been proposed in the literature (see e.g. [7, 23, 51, 73, 75, 79]).

However, instead of reducing the number of components afterwards, it would be beneficial to keep the number of components in the GM approximation small. The GGM developed in [P2] is a method that allows to keep the number of components small by relaxing the non-negativity restriction on component weights, and thus removing the need for reduction methods.

Definition 2 (Generalised Gaussian Mixture) *A $n_{\mathbf{x}}$-dimensional random vector $\underline{\mathbf{x}}$ is a N-component generalised Gaussian mixture*

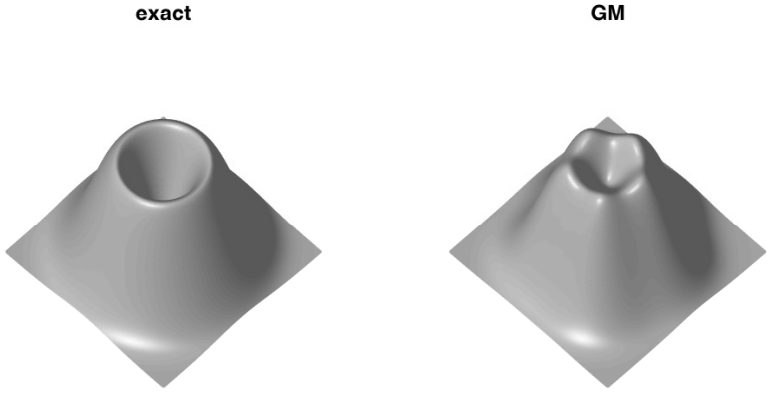


Figure 5: Exact normalised likelihood (same as Figure 4) and its GM approximation with 5 components for a two-dimensional isotropic ranging model.

(GGM) if its probability density function has the form

$$p(\mathbf{x}) = \sum_{j=1}^{N} \alpha_j MVN_{\Sigma_j}^{\mu_j}(\mathbf{x}), \tag{31}$$

where $\alpha_j \in \mathbb{R}$ *and* $\sum_{j=1}^{N} \alpha_j = 1$.

This means that in a GGM some weights can be negative, which is prohibited in standard GMs, but it is required that $p_{\underline{\mathbf{x}}}(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^{n_\mathbf{x}}$.

For the likelihood approximation problem under consideration this relaxation allows to approximate the likelihood in Figure 4 using two Gaussian components. The likelihood of ranging measurement y_k given current estimate $\mathbf{x}$ and transmitter location $\mathbf{c}_k$ is approximated by the GGM as

$$\begin{aligned} p(y_k|\mathbf{x}) &\approx \mathrm{MVN}_{\sigma_{\max}^2\mathbf{I}}^{\mathbf{c}_k}(\mathbf{x}) \cdot \left(1 - \bar{c} \cdot \mathrm{MVN}_{\sigma_{\min}^2\mathbf{I}}^{\mathbf{c}_k}(\mathbf{x})\right) \\ &= \mathrm{MVN}_{\sigma_{\max}^2\mathbf{I}}^{\mathbf{c}_k}(\mathbf{x}) - \bar{c} \cdot \mathrm{MVN}_{\sigma_{\max}^2\mathbf{I}}^{\mathbf{c}_k}(\mathbf{x}) \cdot \mathrm{MVN}_{\sigma_{\min}^2\mathbf{I}}^{\mathbf{c}_k}(\mathbf{x}), \end{aligned} \tag{32}$$

where $\bar{c} = c \cdot (2\pi)^{\frac{n_\mathbf{x}}{2}} {\sigma_{\min}}^{n_\mathbf{x}}$ with $c \leq 1$. (32) is an additive mixture in the form of (31), because the product of two normal PDFs is equal to another scaled normal PDF (see e.g. [5, p. 127] for a proof).

Parameter $\bar{c}$ ensures that (32) is always non-negative and thus is a valid distribution. In addition, $0 < \sigma_{\min} < \sigma_{\max}$ must hold. These two

parameters are defined as

$$\sigma_{\min} = \max\{\epsilon, \beta\, y_k - \sigma\}, \tag{33a}$$

$$\sigma_{\max} = \beta\, y_k + \sigma \tag{33b}$$

with β being a configuration parameter for optimising approximation quality.

The definitions of $\sigma_{\min}$ and $\sigma_{\max}$ are heuristic. However, they are reasonable models as the standard deviation of the GGM components should be proportional to the measurement y_k, because they define where the maximum probability is located, and the slopes of the GGM should resemble the measurement uncertainty σ. The remaining parameter β is determined in [P3] such that the Kullback-Leibler divergence (KLD) between GGM likelihood approximation and exact likelihood is minimised.

Parameter ϵ has to be positive but should be chosen close to zero. It ensures that the GGM also works if $\beta\, y_k - \sigma \leq 0$.

For σ in (33) in [P3] the standard deviations of the zero-mean normal distributions fitted to LOS and NLOS ranging errors are used (see Section 4 for details). Using not necessarily zero-mean normal distributions would introduce one more parameter in (33); the mean of such distribution would have to be added to both $\sigma_{\min}$ and $\sigma_{\max}$.

Figure 6 shows the exact likelihood from Figure 4 and its GGM approximation. Because the GGM's both components use the "volcano's" centre, i.e. the transmitter location, as mean values the GGM approximation's sides are smooth. Achieving such smoothness using the traditional GM method would require a large number of components. Thus, for the problem of approximating the exact measurement likelihood $p(y_k|\mathbf{x})$ the GGM is the better choice than the GM. However, this results should not be generalised. One drawback of the GGM is that the weights of the components have to be chosen so that their sum is a valid PDF, which might be tedious for GGMs with more than two components. For the traditional GM the only constraint, besides non-negativity of the weights, is that the weights sum up to one.

In [P3] it is shown that the GGM's approximation quality depends on σ. The results support the paper's hypothesis that for increasing σ the KLD between normalised GGM likelihood and normalised

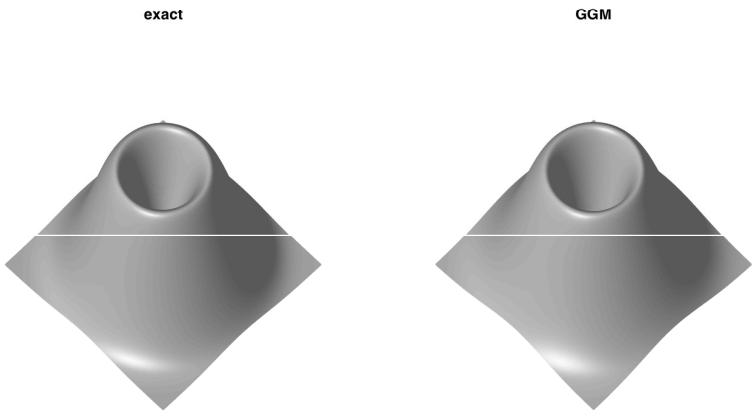


Figure 6: Exact normalised likelihood (same as Figure 4) and its GGM approximation for a two-dimensional isotropic ranging model.

exact likelihood decreases, meaning the approximation quality improves. For $\sigma \to 0$, on the other hand, the GGM fails to model the steep probability gradients of the exact likelihood. This finding provided motivation for research on statistical trilateration with skew-t distributed errors in [P4] and [R1].

The GGM is used for positioning in an UWB network in [P3]. In any unknown UE location each ranging measurement is modelled by the GGM defined in (32), and the product of those GGMs is multiplied with the normal prior for the UE position estimate to obtain the posterior GGM. In [P2] it is proved that mean and covariance of a GGM are computed in the same way as for conventional GMs. Applying those equations to the posterior GGM yields the final UE position estimate (GGM mean) and its uncertainty measure (GGM covariance).

Generalised Gaussian mixture filter

For filtering the GGM filter (GGMF) is used in [P2] and [P3]. Development of the filter is simple. At step $t-1$ the posterior is a GGM in the form of (31), whose mean and covariance are

$$\mathrm{E}(\mathbf{x}) = \sum_{j=1}^{N} \alpha_j \boldsymbol{\mu}_j \triangleq \boldsymbol{\mu}, \tag{34a}$$

$$\mathrm{V}(\mathbf{x}) = \sum_{j=1}^{N} \alpha_j \left(\boldsymbol{\Sigma}_j + (\boldsymbol{\mu}_j - \boldsymbol{\mu})(\boldsymbol{\mu}_j - \boldsymbol{\mu})^T \right) \triangleq \mathbf{P}. \tag{34b}$$

By assuming the posterior at $t-1$ to be normal distributed with mean $\mathrm{E}(\mathbf{x})$ and covariance $\mathrm{V}(\mathbf{x})$ Kalman filtering can be applied for computing a normal prior for step t. In [P2] and [P3] a version of state transition function (2a) is used, namely

$$\underline{\mathbf{x}}_t = \mathbf{F}_{t-1}\underline{\mathbf{x}}_{t-1} + \underline{\mathbf{w}}_{t-1}, \tag{35}$$

where $\mathbf{F}_{t-1}$ is a linear function. The prior is a multivariate normal with mean and variance

$$\hat{\mathbf{x}}_t^- = \mathbf{F}_{t-1}\boldsymbol{\mu}, \tag{36a}$$

$$\mathbf{P}_t^- = \mathbf{F}_{t-1}\mathbf{P}\mathbf{F}_{t-1}^T + \mathbf{Q}_{t-1}. \tag{36b}$$

Modelling the posterior in the GGMF by one Gaussian has the advantage that it keeps the number of Gaussian components low. On the downside, however, information might get lost by this simplification.

Positioning tests with GGM and GGMF

In [P3] the GGMF is compared with the EKF in both simulations and real-world tests for two- and three-dimensional positioning. The GGMF yields better accuracy (lower mean positioning error) and precision (lower median error and 95th percentile error) than the EKF in all six simulation scenarios. Those scenarios simulate two-dimensional and three-dimensional positioning in environments with LOS only, LOS and NLOS, and NLOS only measurements. Also in the real-world tests the GGMF provides significantly better positioning accuracy than the EKF. Especially for three-dimensional positioning the improvement is large. The tests in a real UWB network furthermore show that the GGMF strength is the starting phase of the tracking. It provides precise position estimates from the first location of the test tracks, whereas the EKF needs several steps until it yields satisfying position estimates.

For the tests in [P3] also UKF, a GM filter and a particle filter have been implemented. However, the GM filter has been dropped from consideration because its requirement of large number of components for adequately capturing the likelihood functions prevent it

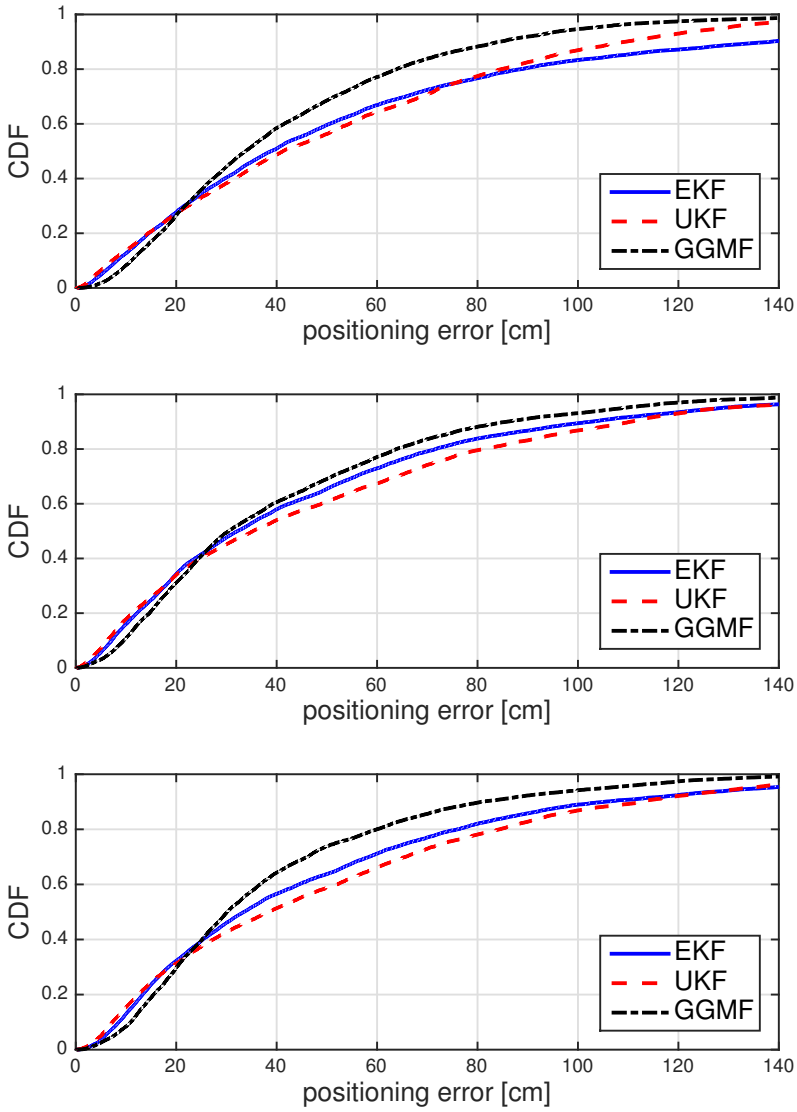


Figure 7: Empirical CDFs of positioning errors from EKF, UKF and GGMF for two-dimensional positioning scenarios in [P3]. Upper plot shows results for scenario in which only LOS measurements are used. Plot in the middle shows results for scenario in which LOS and NLOS measurements are used. Lower plot shows results for scenario in which only NLOS measurements are used.

from being a reasonable choice in terms of computation time. The same holds for the particle filter. With just 100 particles its computation time is approximately 60 times higher than that of the EKF, while its positioning accuracy is unsatisfying. The UKF would be a suitable alternative. Figure 7 shows the empirical positioning error CDFs for UKF in comparison to those of EKF and GGMF for the two-dimensional simulation scenarios from [P3]. For each scenario 100 tracks with 100 time steps each are evaluated using the three filters. In all three scenarios the GGMF is the most precise algorithm (the positioning error CDF is a measure of precision [55]), this means it is more robust against errors in range measurements. The UKF performs very similar to the EKF, and its results are therefore not presented in [P3].

Only for around 40% of the cases do the EKF and UKF provide slightly better position accuracy. This finding is in line with the results of the tests with real-world UWB data in [P3], where the EKF performed on a similar level as the GGMF once it was close to the true track. The reason for this behaviour is that the EKF underestimates the true covariance matrix due to the nonlinearities in the measurement model [8]. Hence, it trusts its prior more than the GGMF does.

The GGM's computational complexity is also analysed and compared with that of the EKF for one time step in [P3]. Its complexity is exponential in the number of ranging measurements available at a specific time step, whereas the EKF's complexity is quadratic in the number of measurements. [P3] gives a more detailed analysis based on five different groups of operations.

For positioning in a mobile phone network in [P2] a different "prior" is used than in [P3]. Instead of using the normal describing the previous time state's posterior that is propagated through the motion model, the posterior of the CA approach explained in Subsection 5.2 is used as prior. The solution of the CA method is also used as prior in the static positioning problems. Furthermore, measurements are RSS values and ranging estimates have to be derived using PL model (21). For real-world applications PL model parameters from the RM would be used, but in the publication's simulations those parameters are assumed to be known.

In [P2] the GGM's positioning accuracy is compared with the CA approach and the EKF update step. Two scenarios are analysed. In the first scenario a poor network geometry is simulated, and at each location signals from at most one transmitter, here called a BS, are received; in the second scenario a good network geometry is simulated, and at each location signals from up to six BSs are received. In both scenarios the GGM outperforms the two reference methods in terms of mean, median and 95th percentile positioning error. Its computation time is for scenario 1 approximately two times and for scenario 2 around 15 times that of CA approach and EKF update step.

For the same two scenarios GGMF and filtered version of the CA approach are compared with the EKF. The filtering is done using a KF, made possible by the linear motion model with additive zero-mean noise. As for the static positioning tests, the GGMF outperforms the two reference methods. Its computation time is for scenario 1 approximately two times and for scenario 2 around 16 times that of filtered CA approach and EKF.

In [P1] GGM and GGMF perform similarly for all four tests in WLANs, meaning that reducing the number of available measurements, low AP density and outdated radio map have only small influence on the achievable accuracy level. Furthermore, in the tests the performance gain of filtering is insignificant for the GGM.

5.4 Descending Gauss-Newton

In the conclusions of [P3] it is mentioned that further work should consider how the GGM concept could be modified to scenarios in which ranging measurements in general are precise but also large ranging errors are possible. One approach that is mentioned in the paper would be to replace the Gaussian components by t-distributed components in the mixture. However, preliminary analysis showed that this would be a rather computational demanding approach.

Therefore, a different approach is considered in [P4] and [R1]. The ranging error $\underline{v}$ in (29) is assumed to be skew-t distributed, which results in larger probabilities for small ranging errors, and enables modelling asymmetries and large deviations in the ranging error distribution. For solving the statistical trilateration an EM approach

is developed in [P4].

The EM approach uses the descending Gauss-Newton (dGN) method, which is a standard nonlinear least squares (NLS) solver. Furthermore, the dGN is used in both [P4] and [R1] as reference positioning method. A detailed explanation of the algorithm can be found, for example, in [69]. Here only a short description is given.

Assuming a multivariate-normal prior $\underline{\mathbf{x}} \sim \mathrm{MVN}(\mathbf{m}, \mathbf{P})$ with position estimate $\mathbf{m}$ and symmetric positive definite (SPD) covariance matrix $\mathbf{P}$ and normal conditional K-dimensional vector $\underline{\mathbf{y}}|(\underline{\mathbf{x}} = \mathbf{x}) \sim \mathrm{MVN}(\mathbf{h}(\mathbf{x}), \mathbf{R})$ of ranging measurements with SPD $\mathbf{R}$, using Bayes' rule (1) the posterior density is

$$\begin{aligned} p_{\underline{\mathbf{x}}|\underline{\mathbf{y}}}(\mathbf{x}|\mathbf{y}) &\propto p_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\mathbf{y}|\mathbf{x}) p_{\underline{\mathbf{x}}}(\mathbf{x}) \\ &\propto \mathrm{e}^{-\frac{1}{2}(\mathbf{h}(\mathbf{x})-\mathbf{y})^T \mathbf{R}^{-1}(\mathbf{h}(\mathbf{x})-\mathbf{y})} \mathrm{e}^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \mathbf{P}^{-1}(\mathbf{x}-\mathbf{m})} \\ &= \mathrm{e}^{-f(\mathbf{x})} \end{aligned} \tag{37}$$

where

$$f(\mathbf{x}) = \frac{1}{2}\left((\mathbf{h}(\mathbf{x})-\mathbf{y})^T \mathbf{R}^{-1}(\mathbf{h}(\mathbf{x})-\mathbf{y}) + (\mathbf{x}-\mathbf{m})^T \mathbf{P}^{-1}(\mathbf{x}-\mathbf{m})\right) \tag{38}$$

is called cost function.

The aim of the dGN is to compute the maximum a-posteriori position estimate $\hat{\mathbf{x}}$, i.e. the mode rather than the mean of the posterior distribution.

Because the measurements in [P4] and [R1] are ranges, $\mathbf{h}(\mathbf{x})$ is a vector of Euclidean distances between transmitters and UE. In the dGN ranging errors have to be assumed to be additive zero-mean normal. Thus, if the normal fitted to ranging error data has a non-zero mean m then this mean is simply subtracted from each ranging measurement y_k to obtain a modified "range" $\tilde{y}_k$ for which the zero-mean assumption holds. Furthermore, measurements are assumed to be mutually independent in [P4] and [R1], resulting in $\mathbf{R}$ being a diagonal matrix with diagonal elements σ^2, which is the variance of the normal fitted to the ranging errors, and thus SPD.

The dGN for statistical trilateration is shown in Algorithm 2. Parameter ϵ_{dGN} should be chosen with care. For example, for the test with

Input: $\mathbf{c}_{1:K}$, $\mathbf{y}$, $\mathbf{m}$, $\mathbf{P}$, m, σ, ϵ_{dGN} // ϵ_{dGN} is threshold value
$\hat{\mathbf{x}} \leftarrow \mathbf{m}$ // initialization
$\tilde{\mathbf{y}} \leftarrow \mathbf{y} - m\mathbf{e}_{K\times 1}$ // modified "data"
$\mathbf{R} \leftarrow \sigma^2 \mathbf{I}_{K\times K}$
$u_{\text{dGN}} \leftarrow \epsilon_{\text{dGN}}$
while $u_{dGN} \geq \epsilon_{dGN}$ **do**
  Compute $\mathbf{H} \leftarrow \frac{\partial \mathbf{h}_{1:K}}{\partial \mathbf{x}}(\hat{\mathbf{x}})$
  if $\mathbf{P}^{-1} \rightarrow \mathbf{0}$ **then**
    $\mathbf{d}_{\text{dGN}} \leftarrow \left(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}\left(\mathbf{h}(\hat{\mathbf{x}}) - \tilde{\mathbf{y}}\right)$
  else
    Compute $\mathbf{K} \leftarrow \mathbf{P}\mathbf{H}^T\left(\mathbf{R} + \mathbf{H}\mathbf{P}\mathbf{H}^T\right)^{-1}$ and
     $\mathbf{d}_{\text{dGN}} \leftarrow \mathbf{m} - \hat{\mathbf{x}} + \mathbf{K}\left(\tilde{\mathbf{y}} - \mathbf{h}(\hat{\mathbf{x}}) - \mathbf{H}\mathbf{m} + \mathbf{H}\hat{\mathbf{x}}\right)$
  end
  Compute $f(\hat{\mathbf{x}})$ using (38) with $\mathbf{x} \leftarrow \hat{\mathbf{x}}$
  Set $\alpha \leftarrow 1$ and compute $f(\hat{\mathbf{x}} + \alpha\mathbf{d}_{\text{dGN}})$ using (38) with
   $\mathbf{x} \leftarrow \hat{\mathbf{x}} + \alpha\mathbf{d}_{\text{dGN}}$
  while $f(\hat{\mathbf{x}} + \alpha\mathbf{d}_{dGN}) \geq f(\hat{\mathbf{x}})$ **do**
    Set $\alpha \leftarrow \alpha/2$, and compute $f(\hat{\mathbf{x}} + \alpha\mathbf{d}_{\text{dGN}})$ using (38) with
     $\mathbf{x} \leftarrow \hat{\mathbf{x}} + \alpha\mathbf{d}_{\text{dGN}}$
  end
  Compute $u_{\text{dGN}} \leftarrow ||\alpha\mathbf{d}_{\text{dGN}}||$ // Euclidean distance of
     consecutive position estimates
  Set $\hat{\mathbf{x}} \leftarrow \hat{\mathbf{x}} + \alpha\mathbf{d}_{\text{dGN}}$
end

Algorithm 2: descending Gauss-Newton algorithm

realistic emulated LTE data in [R1] a value of 1 meter is used. This means, that if the position estimates of two consecutive iterations are closer than 1 meter apart ($u_{\text{dGN}} < 1$ m) the dGN is terminated and the current $\hat{\mathbf{x}}$ is returned. In [P4] and in the numerical experiment in [R1] a fixed number of iterations is used instead, meaning that the outer while loop in Algorithm 2 is replaced by a for loop. For the test with emulated LTE data in [R1] in addition an upper limit for iterations of the dGN is set to avoid an endless loop. The inner while loop, which runs as long as $f(\hat{\mathbf{x}} + \alpha\mathbf{d}_{\text{dGN}}) \geq f(\hat{\mathbf{x}})$, is the reason why the dGN is called "descending". It ensures that the cost function (38) does not increase, which is possible for the standard Gauss-Newton algorithm. For the inner while loop an upper limit for number of iterations should be set, too; in [P4] and [R1] a limit of 5 is used [69].

More sophisticated methods such as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) or the Levenberg-Marquardt algorithm could be used instead of the dGN. However, the dGN was chosen because of its simplicity. For example, the BFGS contains several tuning parameters and will be hard to implement[3]. Thus, the dGN is a good compromise. It offers satisfying positioning accuracy, is simple to implement and computationally fast.

5.5 Expectation Maximization

In [P4] an Expectation Maximization algorithm for solving statistical trilateration problems with additive skew-t distributed measurement errors is developed. The EM computes the MAP position estimate $\hat{\mathbf{x}}$.

The measurement model is $\underline{y}_k|(\underline{\mathbf{x}}=\mathbf{x}) \sim \mathrm{ST}\big(\xi + h_k(\mathbf{x}), \sigma^2, \lambda, \nu\big)$, and posterior pdf (18) is

$$p_{\underline{\mathbf{x}}|\underline{\mathbf{y}}}(\mathbf{x}|\mathbf{y}) \propto p_{\underline{\mathbf{x}}}(\mathbf{x}) \prod_{k=1}^{K} \frac{2}{\sigma} t_\nu\left(\frac{\bar{y}_k}{\sigma}\right) T_{\nu+1}\left(\lambda \frac{\bar{y}_k}{\sigma} \sqrt{\frac{\nu+1}{\nu+\frac{\bar{y}_k^2}{\sigma^2}}}\right), \tag{39}$$

where $\bar{y}_k = y_k - h_k(\mathbf{x}) - \xi$. The measurements are ranges defined as Euclidean distances between transmitter and receiver and are assumed to be mutually independent. Furthermore, the prior is assumed to be a multivariate-normal $\underline{\mathbf{x}} \sim \mathrm{MVN}(\mathbf{m}, \mathbf{P})$.

Instead of the skew-t distributed measurement model the hierarchical version

$$\underline{y}_k|(\underline{\mathbf{x}}=\mathbf{x}, \underline{t}_k = t_k, \underline{\tau}_k = \tau_k) \sim \mathrm{N}\left(\xi + h_k(\mathbf{x}) + \delta_\lambda t_k, \frac{1-\delta_\lambda^2}{\tau_k}\sigma^2\right), \tag{40}$$

with hyperparameters $\underline{t}_k|(\underline{\tau}_k = \tau_k) = |\sigma \underline{w}_k/\sqrt{\tau_k}|$, where $\underline{w}_k \sim \mathrm{N}(0,1)$, and $\underline{\tau}_k \sim \mathrm{Gamma}(\nu/2, \nu/2)$ is used.

The idea behind the EM algorithm developed in [P4] is to iteratively update the position estimate $\hat{\mathbf{x}}$ with fixed hyperparameters in the maximization step (M-step) and update the hyperparameters with fixed $\hat{\mathbf{x}}$ in the expectation step (E-step). Since for fixed hyperparameters the measurement errors are normal distributed according to

[3] BFGS implementations exist for mathematical programs/ programming languages.

(40) the dGN can be used in the M-step. The hyperparameters are updated using their conditional distributions.

Given the K independent scalar measurements $\mathbf{y} = y_{1:K}$, the dGN computes for the j-th iteration the mode of the conditional posterior $p(\mathbf{x}^{(j)}|\mathbf{y}, t_{1:K}^{(j-1)}, \tau_{1:K}^{(j-1)})$, that is, it solves the minimization problem

$$\hat{\mathbf{x}} \leftarrow \arg\min_{\mathbf{x}} \Big((\mathbf{x}-\mathbf{m})^T \mathbf{P}^{-1} (\mathbf{x}-\mathbf{m}) + \frac{\sum_{k=1}^{K} \tau_k^{(j-1)} \left(y_k - h_k(\mathbf{x}) - \xi - \delta_\lambda t_k^{(j-1)} \right)^2}{\sigma^2 \left(1-\delta_\lambda^2\right)} \Big), \tag{41}$$

where $\delta_\lambda = \lambda/\sqrt{1+\lambda^2}$. The hyperparameters are initilaized as $t_k \leftarrow -\xi/\delta_\lambda$ and $\tau_k \leftarrow 1$ to ensure that the dGN finds the minimizer of the likelihood for normal noise.

The conditional distribution of $\underline{\tau}_k$, that is $p(\tau_k|\mathbf{x}, \mathbf{y}_{1:K}, t_{1:K}, \tau_{-k})$ with $\tau_{-k} = \{\tau_1, .., \tau_{k-1}, \tau_{k+1}, .., \tau_K\}$, is a Gamma distribution with shape $\nu/2$ and and scale parameter

$$\frac{2\sigma^2 \left(1-\delta_\lambda^2\right) \nu}{\left(y_k - h_k(\mathbf{x}^{(j)}) - \xi - \delta_\lambda t_k^{(j-1)} \right)^2 + 4\sigma^2 \left(1-\delta_\lambda^2\right)}.$$

The product of shape and scale is the distribution's mean, which is used as updated τ_k.

For $\underline{t}_k$ its conditional distribution $p(t_k|\mathbf{x}, \mathbf{y}_{1:K}, t_{-k}, \tau_{1:K})$, where $t_{-k} = \{t_1, .., t_{k-1}, t_{k+1}, .., t_K\}$, is a truncated normal distribution bounded from below at zero, i.e. $t_k \geq 0$ has to hold. The truncated normal distribution's center is $\mu_k = \frac{1}{2\delta_\lambda} \left(y_k - h_k(\mathbf{x}) - \xi \right)$ and its scale is $\sigma_k = \sqrt{\left(\sigma^2 \left(1-\delta_\lambda^2\right)\right)/(2\tau_k)}$. Similarly to τ_k, its mean value $\mu_k + \sigma_k \phi(-\mu_k/\sigma_k)/\left(1-\Phi(-\mu_k/\sigma_k)\right)$ is used as updated t_k. For large negative values of μ_k/σ_k this equation for the mean gives significant numerical floating point computation errors. Therefore, in [R1] the mean is computed using the scaled complementary error erfcx, which is defined as $\mathrm{erfcx}(x) = \mathrm{e}^{x^2} \cdot \mathrm{erfc}(x)$; $\phi(-\mu_k/\sigma_k)/\left(1-\Phi(-\mu_k/\sigma_k)\right)$ in the mean equation is replaced by $\sqrt{\frac{2}{\pi}}/\mathrm{erfcx}\left(\frac{-\mu_k/\sigma_k}{\sqrt{2}}\right)$.

Input: $\mathbf{c}_{1:K}$, $\mathbf{y}$, $\mathbf{m}$, $\mathbf{P}$, m, σ, ϵ_{EM}, ϵ_{dGN} // ϵ_{EM} and ϵ_{dGN} are threshold value

$\mathbf{x}^{(0)} \leftarrow \mathbf{m}$, $t_{1:K}^{(0)} \leftarrow -\xi/\delta_\lambda$ and $\tau_{1:K}^{(0)} \leftarrow 1$ // initialization

$u_{\mathrm{EM}} \leftarrow \epsilon_{\mathrm{EM}}$

while $u_{EM} \geq \epsilon_{EM}$ **do**

  Initialize $\hat{\mathbf{x}} \leftarrow \mathbf{x}^{(j-1)}$, compute $\tilde{\mathbf{y}} \leftarrow \mathbf{y} - \xi - \delta_\lambda t_{1:K}^{(j-1)}$ and $\mathbf{R} \leftarrow \sigma^2(1-\delta_\lambda^2)\mathrm{diag}\{1/\tau_1^{(j-1)}, .., 1/\tau_K^{(j-1)}\}$

  $u_{\mathrm{dGN}} \leftarrow \epsilon_{\mathrm{dGN}}$

  while $u_{dGN} \geq \epsilon_{dGN}$ **do**

    // Run outer while loop of dGN as explained in Algorithm 2

  end

  Set $\mathbf{x}^{(j)} \leftarrow \hat{\mathbf{x}}$

  Compute $t_{1:K}^{(j)}$ and $\tau_{1:K}^{(j)}$ as $\tau_k^{(j)} \leftarrow \frac{\sigma^2(1-\delta_\lambda^2)\nu^2}{\left(y_k - h_k(\mathbf{x}^{(j)}) - \xi - \delta_\lambda t_k^{(j-1)}\right)^2 + 4\sigma^2(1-\delta_\lambda^2)}$ and $t_k^{(j)} \leftarrow \mu_k + \sigma_k\sqrt{\frac{2}{\pi}}/\mathrm{erfcx}\left(\frac{-\mu_k/\sigma_k}{\sqrt{2}}\right)$ where $\mathrm{erfcx} \leftarrow \mathrm{e}^{x^2}\mathrm{erfc}(x)$ is the scaled complementary error function, $\mu_k \leftarrow \frac{1}{2\delta_\lambda}\left(y_k - h_k(\mathbf{x}^{(j)}) - \xi\right)$ and $\sigma_k \leftarrow \sqrt{\frac{\sigma^2(1-\delta_\lambda^2)}{2\tau_k^{(j)}}}$

  Compute $u_{\mathrm{EM}} \leftarrow ||\mathbf{x}^{(j)} - \mathbf{x}^{(j-1)}||$

end

Algorithm 3: Expectation Maximization algorithm

The EM for statistical trilateration is shown in Algorithm 3. Parameters ϵ_{EM} and ϵ_{dGN} are set in [R1] to 1 meter for the test with realistic emulated LTE data. This means, that if the position estimates of two consecutive iterations in either EM or dGN are closer than 1 meter apart then the corresponding algorithm is terminated and its current MAP estimate is returned. In [P4] and in the numerical experiment in [R1] fixed number of iterations, n_{EM} for the EM and n_{dGN} for the dGN, are used instead, meaning that the while loops are replaced by for loops in Algorithm 3. For the test with emulated LTE data in [R1] in addition upper limits of $n_{\mathrm{EM}} = 50$ and $n_{\mathrm{dGN}} = 50$ are used to avoid endless while loops. Furthermore, the number of iterations to find α in the dGN is limited to 5 [69] in [P4] and [R1].

Computational complexity of EM and dGN

The computational complexity of EM and dGN is analysed in both [P4] and [R1]. In [P4] it is shown that the number of iterations used by EM n_{EM} and dGN n_{dGN} specify the two algorithms' computational requirements, and that the EM has n_{EM} times higher computational complexity than the dGN for the five operation classes addition, subtraction, multiplication, division, and other. Furthermore, the analysis shows that the E-step in the EM algorithm is computationally light compared with the M-step.

In [R1] it is analysed how many iterations both EM and dGN use when Algorithms 2 and 3 are used, i.e. while loops are used and n_{EM} and n_{dGN} only act as upper limits to avoid endless loops. For the test with emulated LTE data, the dGN uses on average 4 to 6 iterations. The EM uses on average 5 to 8 iterations, and the dGN run inside the EM uses on average 2 to 3 iterations.

Positioning accuracy of EM and dGN

In the numerical experiment of [P4] the EM clearly outperforms the standalone dGN; it reduces mean and median positioning errors by approximately 25% and the 95th percentile of all errors by approximately 13%.

The numerical experiment is repeated in [R1] under the assumption that ranging errors are skew-t distributed with parameters fitted by a GS to ranging error data from the two LTE channel models EPA5 and ETU70. The experiments support the finding of [P4]. Reductions of 4% to 39% in the error quantiles are observed.

However, the test with emulated LTE data in [R1] draws a different picture. While the EM outperforms the dGN for the EPA5 channel, the dGN performs better for the ETU70 channel. The analysis in the paper shows that the EM's positioning accuracy is more strongly influenced by the signal-to-noise ratio (SNR) than the dGN. For the test parameters of skew-t and normal distributions are fitted to training data with high SNR (offline phase), but in the online phase positioning is done in locations with low, medium and high SNR. If the SNR is high the EM outperforms the dGN, for medium SNR both methods perform similarly, and for low SNR the dGN has better positioning

accuracy than the EM. The online data of the ETU70 channel is dominated by medium and low SNR values, while for the EPA5 channel medium SNR values dominate. For decreasing SNR the error distributions for both channel models change significantly. The slopes at both sides of the error distribution's peak get flatter, and the errors resemble more closely a normal distribution. This explains why the EM performs better than the dGN for EPA5 test, and the dGN performs better for ETU70 test.

In [R1] it is concluded that for high SNR training data the EM algorithm is unable to correctly refine the channel parameters for low SNR scenarios in the online phase, and that the EM should outperform the dGN if SNR values would be similar in training and positioning data. Thus, the ability of modelling ranging errors more precisely (by skew-t distribution) and providing better positioning accuracy than the dGN comes at the cost of higher vulnerability to different error distributions in training and positioning data. However, being able to have a more realistic model for the measurement errors was the reason for using a skew-t distributed error model in the first place, so this vulnerability should not be interpreted as weakness. It rather shows that the skew-t model in fact is more realistic than the normal model for the tested data.

5.6 Gibbs sampler

When one is interested in the posterior distribution's mean rather than its mode, the GS [31] can be used to solve the trilateration problem. The GS developed here has not been published but is mentioned briefly in [R1]. For comparison it is presented in this thesis.

As for the EM algorithm in the previous subsection, the measurement model is $\underline{y}_k|(\underline{\mathbf{x}} = \mathbf{x}) \sim \mathrm{ST}\big(\xi + h_k(\mathbf{x}), \sigma^2, \lambda, \nu\big)$ and the posterior pdf is (39). Instead of the skew-t distributed measurement model its hierarchical version (40) is used.

The derivation of the GS for positioning follows closely the derivation of the EM in [P4]. Assuming a multivariate-normal prior distribution for $\underline{\mathbf{x}}$ with mean $\mathbf{m}$ and covariance $\mathbf{P}$ and given the priors for the hyperparameters of (40), the joint prior distribution is

$$p_{\underline{\mathbf{x}},\underline{t}_{1:K},\underline{\tau}_{1:K}}(\mathbf{x}, t_{1:K}, \tau_{1:K}) \quad \propto \quad p_{\underline{\mathbf{x}}}(\mathbf{x}) \cdot p_{\underline{t}_{1:K}|\underline{\tau}_{1:K}}(t_{1:K}|\tau_{1:K}) \cdot p_{\underline{\tau}_{1:K}}(\tau_{1:K})$$

$$= \quad e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \mathbf{P}^{-1}(\mathbf{x}-\mathbf{m})} \cdot \prod_{k=1}^{K} [t_k \geq 0]$$
$$\cdot e^{-\frac{1}{2\sigma^2}\sum_{k=1}^{K} \tau_k t_k^2} \cdot \prod_{k=1}^{K} \tau_k^{\frac{\nu}{2}-1} e^{-\frac{2}{\nu}\tau_k}. \qquad (42)$$

Thus, the joint posterior given the K independent scalar measurements $\mathbf{y}$ is

$$p(\mathbf{x}, t_{1:K}, \tau_{1:K}|\mathbf{y}) \propto e^{-\frac{1}{2\sigma^2\left(1-\delta_\lambda^2\right)}\sum_{k=1}^{K} \tau_k \left(\bar{y}_k - \delta_\lambda t_k\right)^2} \cdot e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \mathbf{P}^{-1}(\mathbf{x}-\mathbf{m})}$$
$$\cdot \prod_{k=1}^{K} [t_k \geq 0] e^{-\frac{1}{2\sigma^2}\sum_{k=1}^{K} \tau_k t_k^2} \cdot \prod_{k=1}^{K} \tau_k^{\frac{\nu}{2}-1} e^{-\frac{2}{\nu}\tau_k} \qquad (43)$$

As mentioned in Section 4 the idea of the GS is to sample from conditional distributions if sampling from the posterior (43) is unfeasible. Therefore, the GS tries to sample from the conditional distribution of $\underline{\mathbf{x}}$, which is

$$p(\mathbf{x}|\mathbf{y}, t_{1:K}, \tau_{1:K}) \propto e^{-\frac{\sum_{k=1}^{K} \tau_k (y_k - h_k(\mathbf{x}) - \xi - \delta_\lambda t_k)^2}{2\sigma^2\left(1-\delta_\lambda^2\right)}} \cdot e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \mathbf{P}^{-1}(\mathbf{x}-\mathbf{m})}. \qquad (44)$$

Because of $h_k(\cdot)$ being a nonlinear function, the GS cannot be directly applied. Instead $h_k(\cdot)$ has to be replaced by its linear approximation. This means that samples of the Gibbs sampler with linear approximation will not be from the true posterior.

Let $h_k(\mathbf{x}) \approx h_k(\mathbf{x}^*) + \mathbf{H}_k(\mathbf{x} - \mathbf{x}^*)$, where $\mathbf{H}_k$ is the derivative of $h_k(\cdot)$ evaluated at $\mathbf{x}^*$. Then the conditional distribution of $\underline{\mathbf{x}}$ is

$$p(\mathbf{x}|\mathbf{y}, t_{1:K}, \tau_{1:K}) \propto e^{-\frac{\sum_{k=1}^{K} \tau_k (y_k - h_k(\mathbf{x}) - \xi - \delta_\lambda t_k)^2}{2\sigma^2\left(1-\delta_\lambda^2\right)}} \cdot e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \mathbf{P}^{-1}(\mathbf{x}-\mathbf{m})} \qquad (45)$$
$$\approx e^{-\frac{\sum_{k=1}^{K} \tau_k (\tilde{y}_k - \mathbf{H}_k \mathbf{x})^2}{2\sigma^2\left(1-\delta_\lambda^2\right)}} \cdot e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T \mathbf{P}^{-1}(\mathbf{x}-\mathbf{m})},$$

where $\tilde{y}_k = y_k - \xi - \delta_\lambda t_k - h_k(\mathbf{x}^*) + \mathbf{H}_k \mathbf{x}^*$. By examination of (45), the conditional distribution from which the jth sample $\mathbf{x}^{(j)}$ is drawn in the GS is

$$\underline{\mathbf{x}}^{(j)} \sim \mathrm{MVN}(\mathbf{m}_K, \mathbf{P}_K), \qquad (46)$$

which is a multivariate normal distribution with $\mathbf{m}_K$ and $\mathbf{P}_K$ computed recursively using $\mathbf{m}$ and $\mathbf{P}$.

The conditional distributions for $\underline{t}_k$ and $\underline{\tau}_k$ are the same as for the EM. Thus, for $\underline{t}_k$ its conditional distribution is a truncated normal distribution bounded from below at zero, and for $\underline{\tau}_k$ it is a Gamma distribution.

Input: $\mathbf{c}_{1:K}$, $\mathbf{y}$, $\mathbf{m}$, $\mathbf{P}$, T_0, T // T_0 is number of "burn-in" samples and T is number of retained samples
$t_{1:K}^{(0)} \leftarrow -\xi/\delta_\lambda$ and $\tau_{1:K}^{(0)} \leftarrow 1$ // initialization
Draw $\mathbf{x}^{(0)}$ from $\underline{\mathbf{x}} \sim \mathrm{MVN}(\mathbf{m}, \mathbf{P})$
Set $\mathbf{m}_0 \leftarrow \mathbf{m}$ and $\mathbf{P}_0 \leftarrow \mathbf{P}$
for $j = 1$ *to* $T_0 + T$ **do**
  Compute $\mathbf{H} \leftarrow \frac{\partial \mathbf{h}_{1:K}}{\partial \mathbf{x}}\left(\mathbf{x}^{(j-1)}\right)$
  for $k = 1$ *to* K **do**
    $S_k \leftarrow \mathbf{H}_k \mathbf{P}_{k-1} \mathbf{H}_k^T + \frac{\sigma^2(1-\delta_\lambda^2)}{\tau_k}$
    $K_k \leftarrow \mathbf{P}_{k-1} \mathbf{H}_k^T S_k^{-1}$
    $\mathbf{m}_k \leftarrow \mathbf{m}_{k-1} + K_k\left(\tilde{y}_k - \mathbf{H}_k \mathbf{m}_{k-1}\right)$
    with $\tilde{y}_k \leftarrow y_k - \xi - \delta_\lambda t_k - h_k(\mathbf{x}^{(j-1)}) + \mathbf{H}_k \mathbf{x}^{(j-1)}$
    $\mathbf{P}_k \leftarrow \mathbf{P}_{k-1} - K_k S_k K_k^T$
  end
  Draw $\mathbf{x}^{(j)}$ from $\underline{\mathbf{x}} \sim \mathrm{MVN}(\mathbf{m}_K, \mathbf{P}_K)$
  Draw $\tau_{1:K}^{(j)}$ from $\underline{\tau}_k \sim \mathrm{Gamma}\left(\frac{\nu}{2}, \frac{2\sigma^2(1-\delta_\lambda^2)\nu}{\left(y_k - h_k(\mathbf{x}^{(j)}) - \xi - \delta_\lambda t_k^{(j-1)}\right)^2 + 4\sigma^2(1-\delta_\lambda^2)}\right)$
  Draw $t_{1:K}^{(j)}$ from $\underline{t}_k \sim \mathrm{TN}_{\geq 0}\left(\frac{1}{2\delta_\lambda}\left(y_k - h_k(\mathbf{x}^{(j)}) - \xi\right), \frac{\sigma^2(1-\delta_\lambda^2)}{2\tau_k^{(j)}}\right)$
end
Compute posterior mean $\hat{\mathbf{x}}_{\mathrm{pm}} \leftarrow \frac{1}{T}\sum_{j=1}^{T} \mathbf{x}^{(T_0+j)}$

Algorithm 4: Gibbs sampler algorithm

Algorithm 4 shows the GS for statistical trilateration, which yields the posterior mean $\hat{\mathbf{x}}_{\mathrm{pm}}$. Drawing t_k's from the truncated normal distribution is, generally, the most complicated part of the algorithm. Simply drawing from a normal distribution that uses the truncated normal's center μ_k as mean and its scale σ_k as standard deviation until $t_k \geq 0$ is drawn requires an average of $1/\Phi(\mu_k/\sigma_k)$ draws [71, pp. 52 ff.], where $\Phi(.)$ is the cdf of the standard normal distribution. Thus, for large negative values of μ_k/σ_k this approach is unfeasible. In [71] the authors suggest an alternative method that uses a translated exponential distribution.

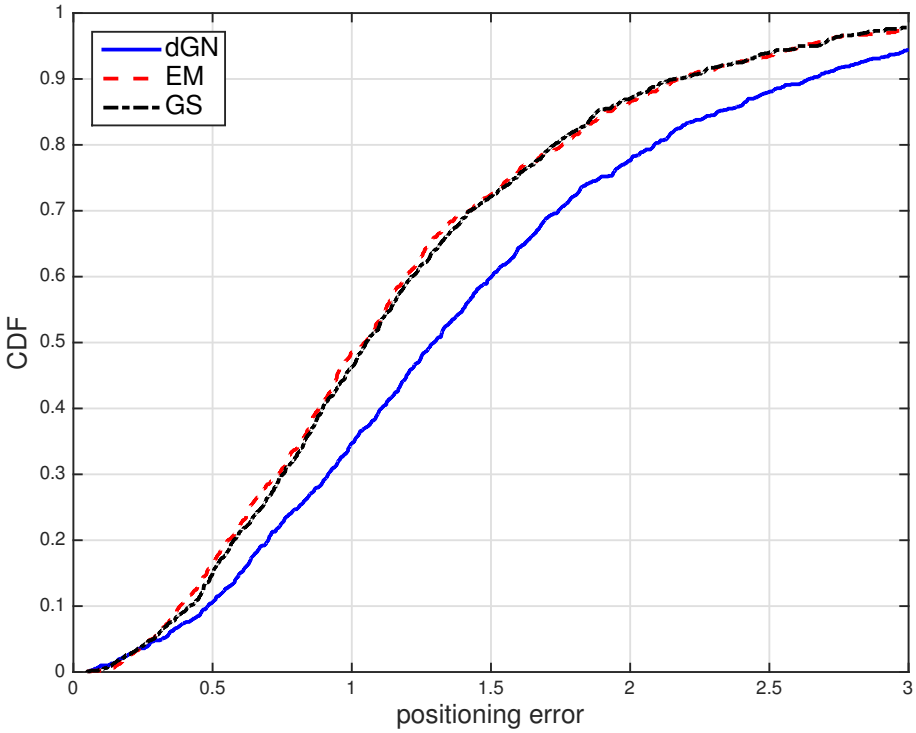


Figure 8: Empirical CDFs of positioning errors for dGN, EM and GS in setup of numerical experiment from [P4].

Positioning accuracy and computational complexity of the GS

The computational complexity of the GS depends on the number of "burn-in" and retained samples. In order to achieve satisfying positioning precision large number of samples are required. For example, in the setup of the numerical experiment from [P4] the GS uses $T_0 = 500$ "burn-in" and $T = 1\,000$ retained samples to achieve a similar precision as the EM algorithm. Using this amount of samples results in an approximately 500 times higher computation time compared with EM and an approximately 2 500 times higher computation time compared with dGN, making it impractical for real-time positioning on mobile handheld devices. The only advantage the GS has in comparison with the EM is that it allows to compute a uncertainty measure for its position estimate.

Figure 8 shows the empirical positioning error CDFs of dGN, EM and GS in the numerical experiment from [P4]. In the experiment the

two-dimensional position $\underline{\mathbf{x}}$ is assumed to have the prior distribution

$$\underline{\mathbf{x}} \sim \mathrm{MVN}\left(\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right], 100\,\mathrm{I}_{2\times 2}\right). \tag{47}$$

Four transmitters are located at the corners of a 40-by-40 square centred at $(0,0)$. Here 1 000 UE positions are drawn from the prior distribution (47). For each position $K = 12$ independent distance measurements (three to each reference node) are drawn from

$$\underline{y}_k \left| \left(\underline{\mathbf{x}} = \mathbf{x}\right) \sim \mathrm{ST}\left(\xi + h_k(\mathbf{x}), \sigma^2, \lambda, \nu\right) = \mathrm{ST}\left(2 + h_k(\mathbf{x}), 3^2, 3, 3\right), \right. \tag{48}$$

where $h_k(\mathbf{x})$ is the Euclidean distance between $\mathbf{x}$ and the transmitter.

The empirical CDFs in Figure 8 show that both EM and GS are significantly more precise, this means more robust [55], than the dGN. Only for the "best" approximately 5% of the 1 000 test locations the dGN can compete with EM and dGN.

5.7 Alternative positioning concepts and systems

Besides the methods presented in this section there exist numerous alternative positioning algorithms that cannot be covered in this thesis. Therefore, only a small fraction of them are briefly discussed in this subsection.

In Subsection 2.1 alternative Gaussian filters besides EKF and UKF, non-Gaussian filters and mixture filters are discussed. All of these filters can be used for positioning and navigation.

Subsection 3.4 discusses RM approaches for alternative positioning systems. One of them is to store FPs from geomagnetic fields in the RM and use them for indoor positioning. In [20] the magnetic field in the unknown UE location is measured by an array of e-compasses. To infer the location the root means square difference between the measurement and the RM entries is computed. This means positioning is done using a deterministic nonparametric method, which have been discussed in Subsection 5.1 before. Besides nonparametric method also parametric methods could be used. For example, if coils are deployed at known locations then RSS values of the magnetic field could be used to derive distance measurements. Using these distance the UE position could be inferred by trilateration (see [52]

and references within for details). In [30] the authors suggested to store sequences of magnetic FPs in the RM. For positioning the UE measurement is again compared with the RM. However, instead of comparing only one UE measurement a sequence of measurements is compared with the stored sequences. The reasoning behind the approach is that in indoor environments individuals tend to move on certain tracks while avoiding other areas. An obvious drawback of the approach is that it requires the UE to be moving to measure signal sequences. Static positioning is highly unreliable because magnetic signals are not spatially unique [30].

Another concept are distributed cooperative positioning techniques. Especially in indoor environments NLOS channel conditions can prevent high positioning accuracy [25]. Cooperative positioning tackles this problem by cooperation amongst several UEs. By cooperating the UEs obtain additional knowledge which can be used to compute more precise position estimates. Shared information can be the UEs' position estimates, distance estimates to transmitters (with known locations), etc. More details can be found, for example, in [25] and references therein.

Simultaneous localisation and mapping has been developed in the field of robotics [25, 37]. A SLAM algorithm is, for example, developed in [30] for a geomagnetic field system. In SLAM the UE tries to localise itself and generate a map of its environment simultaneously. For localising range-related measurements with reference points are used. The difference to the triangulation techniques discussed above is that the reference points' locations are unknown and are estimated simultaneously, which is the mapping part of SLAM [25]. For surveys on SLAM the reader is referred to [25] and references therein.

Another approach for positioning is dead reckoning (DR), which can be used for positioning and navigation in indoor environments for which no FP-RMs are available. DR estimates the UE's relative position using sensor measurements. Given a known (or estimated) initial position estimate the aim of DR systems is to compute changes in the UE position over time [37]. In pedestrian DR (PDR) the measurements from IMUs are used for positioning. Measurements can include angular velocity, a combination of gravitational and inertial linear acceleration, heading information (if the IMU has a 3D mag-

netometer) and altitude (if IMU has a barometer) [25]. By temporal integration of accelerometer and gyroscope data from the IMU the UE position can be estimated [26, p. 63].

An extensive survey on inertial PDR can be found in [37]. Besides PDR systems the survey also shows how to fuse them with other systems, which is necessary for reliable long-term navigation and tracking. In the paper using a combination of PDR and map information or PDR and WLAN fingerprinting is discussed, but also other alternatives are mentioned. Generally, fusing various positioning systems into a hybrid system can improve positioning accuracy and precision significantly if the advantages and disadvantages of the underlying systems counter each other [37].

6 Conclusions and future work

6.1 Positioning with nonlinear measurement models

The first question this thesis wanted to answer was "How can the UE position be calculated precisely, reliable, efficiently and with low cost for nonlinear measurement models?" As answers two algorithms, the generalised Gaussian mixture algorithm and the EM algorithm for skew-t distributed errors have been developed.

For mildly nonlinear measurement models NLS (which is here the EKF update step) and EKF yield, in general, satisfactory positioning accuracy, but for highly nonlinear models they can seriously underestimate the posterior covariance. In addition, they are unable to capture multiple peaks of the likelihood function.

Therefore, in [P2] a generalised Gaussian mixture approach, which relaxes the restriction of GM to nonnegative component weights, and its filter is developed. In the publication it is proven that mean and covariance of the GGM are computed with the same equations as for the common GM. The GGM allows to reduce the number of components significantly, compared with the GM, in certain applications. One such application is the modelling of measurement likelihood functions for transmitters with isotropic antennas, where the GGM uses only two components.

In [P2] the GGM's and GGMF's positioning performances are compared with those of the CA approach and its filtered version (CAF), and the NLS and the EKF by simulations for a cellular telephone network. The GGM and GGMF, which use BS-IDs and RSS values, outperform CA and CAF, which use only BS-IDs, and NLS and EKF, which use BS-IDs and RSS values, for both poor and good network geometry.

The GGM is analysed in more detail in [P3] and is slightly modified. The analysis shows how the GGM's approximation quality depends on the measurement error's variance. For high variance it provides precise approximations, but for low variance it fails to do so. Thus, its applications for approximating likelihoods of highly reliable measurements cannot be recommended.

For [P3] simulations and tests with real-world UWB data were performed. Although UWB yields precise range measurements, at least for LOS, the GGMF outperforms the EKF, which is run for comparison, in scenarios for both good and poor network geometry. The real-world tests show that using the GGMF is particularly advantageous in the beginning of a tracking task. In [P1] the GGM and the GGMF are tested and compared with several other parametric and one nonparametric positioning method in a WLAN.

As stated above, the GGM was developed to tackle the problem that the traditional GM requires too many components for modelling the measurement likelihood of an isotropic transmitter's nonlinear measurement model, which causes it to be too slow for real-time positioning on small mobile devices. Another answer for the first research question would be particle filters (aka sequential Monte Carlo methods). However, they suffer from the same drawback as the GM; they are too computationally demanding and too slow for real-time positioning on smartphones.

In [P3] replacing the normal distributions by t-distributions in the generalised mixture framework has been proposed as yet another alternative. However, initial work showed that this approach most likely results in large computational complexity and thus long computation time. Therefore, instead an Expectation Maximisation algorithm for skew-t distributed measurement errors has been proposed in [P4], which also answers the second research question of this thesis.

6.2 Heavy-tailed and skewed measurement errors

The second question this thesis wanted to answer was "How can heavy-tailed and/ or skewed distributed measurement errors be properly accounted for in positioning problems?" As answer an EM algorithm has been developed.

If the measurement error distribution is close to a normal distribution, i.e. only slightly heavy-tailed and slightly skewed, then in case of a nonlinear measurement model a NLS algorithm such as a descending Gauss-Newton algorithm provides satisfying positioning accuracy. However, for significantly heavy-tailed and/ or significantly skewed error distributions the NLS algorithms' accuracies degrade.

These types of error distributions have been detected, for example, for NLOS range measurements from an UWB network and for range measurements from an LTE network.

Therefore, in [P4] an Expectation Maximisation algorithm for solving statistical trilateration problems with skew-t distributed measurement errors is developed. The EM algorithm uses a hierarchical model of the skew-t distributions that allows the usage of an NLS algorithm in the Maximisation step. In the Expectation step the hierarchical model's hyperparameters are updated. The algorithm's novelty is that it accounts for both heavy-tailedness and skewness in the error. For heavy-tailed, unskewed ranging errors and nonlinear measurement function there has been proposed a solution, for example, in [70].

The numerical experiment in [P4] shows that the EM improves the positioning accuracy significantly compared with a standard NLS algorithm, which works under the assumption of normal errors, if ranging errors are indeed skew-t distributed. The analysis of the computational complexity reveals that the improvement in accuracy comes at the cost of increased computation time. This increase is controlled by the number of EM iterations. However, the analysis in [R1] shows that the EM that uses a dGN in its M-step has an only approximately fives times higher computation time than a standalone dGN, which is in line with the findings of [P4]. Thus, the EM is computationally light enough for real-time positioning on smartphones with low computational power.

For fitting the parameters of skew-t distributions to training data in the offline phase a Gibbs sampler is used in [P4] and [R1], which yields posterior distributions for the parameter estimates. The Gibbs sampler could also be used in the online phase as an alternative to NLS and EM methods, which only yield point position estimates and no information on the estimates uncertainty. However, in this thesis it is shown that the GS is too slow for real-time positioning on small mobile devices. In the simulation environment from [P4] it is approximately 500 times slower than the EM and around 2 500 times slower than the dGN, while providing similar positioning accuracy as the EM algorithm. Furthermore, it is important to notice that the Gibbs sampler cannot sample from the true posterior distribution

but only from an approximation.

The EM algorithm is analysed thoroughly in [R1] using two LTE channel models. In the paper parameters of normal and skew-t distributions are fitted to LTE data for EPA5 and ETU70 channel models. Depending on the channel the fitted skew-t distributions model ranging errors much better than the normal distributions. The numerical experiment based on the fitted distributions shows that the EM outperforms a NLS for both channels, even when measurement errors are almost normal distributed. However, in the test with real-world emulated LTE data the EM fails to outperform the NLS, which is used for comparison. The analysis of the results shows that disparities in the SNR between training and actual measurements cause a significant performance degradation for the EM. This means, that the EM requires training data that is very close to the data used in the online phase in order to yield higher positioning accuracy than the NLS.

Instead of the proposed EM algorithm maximum likelihood estimators could also be used for estimating the UE position. There exist a variety of off-the-shelf optimisation codes for computing the maximum likelihood. However, the EM has a couple of advantageous that makes it a good choice for solving the trilateration problem under consideration (see [60, p. 28 f.] for a more thorough discussion). Firstly, the EM is numerically stable. Secondly, the EM is known to converge almost always to a local MAP. Convergence to a global MAP cannot be guaranteed if there are multiple MAPs, but generally all optimisation algorithm suffer from this drawback. Thirdly, the EM is easy to implement. For the E-step closed form equations for updating the hyperparameters of the hierarchical model are derived in [P4], and for the M-step standard statistical methods such as the dGN can be used. Using the hierarchical model and EM furthermore ensures that neither the posterior PDF has to be evaluated nor its derivative has to be computed. Especially the latter task seems to be complicated. Fourthly, the storage requirements for the proposed EM are small. Only the current estimates of state and hyperparameters have to be stored. Fifthly, for the considered trilateration problem the EM showed fast convergence. In the tests in [R1] the EM used less than 11 iterations in 95% of the cases. Sixthly, the EM makes it easy to use prior information on the position estimate.

6.3 Future work

The GGM approach has been developed and analysed for isotropic antennas. It might be also applicable for modelling likelihood functions for directional antennas. However, the proof of this hypothesis and how to do it is left for further research.

Furthermore, it was shown in [P3] that the GGM is ill-suited for highly accurate range measurements. It could be tested if using a GM noise would improve the performance considerably, and how it would perform in comparison with the skew-t distributed noise.

In the previous subsection it is stated that the EM requires training data that is very close to the data used in the online phase in order to yield higher positioning accuracy than the NLS. It is left for future research to develop and test a mathematical model that describes how changes in the SNR change the error distribution. Using such a model should allow the EM to clearly outperform the NLS in the tests with LTE data.

Another open question is how the EM algorithm and a filtered version of it perform in the testbed of [P1]. This paper focuses on parametric fingerprinting methods. It defines how parametric methods work and how they differ from nonparametric methods. In the paper both offline and online phase of several parametric and one nonparametric method are presented. The tests with extensive WLAN data reveal that the parametric methods reduce the RM size significantly compared with the nonparametric method. Furthermore, they show that parametric methods require less measurements and a lower AP density than the nonparametric method to achieve satisfactory positioning accuracy. How to fuse these methods with inertial measurement units and floor map information is left for future research.

The question on how to integrate the EM algorithm in a filter has not been solved yet. Since the EM does not yield a covariance matrix for the position estimate Bayesian filtering cannot be applied. However, since the finalisation of [P1]–[P4] and [R1] an algorithm for solving the filtering problem when measurement errors are skew-t distributed has been proposed in [66], namely a recursive skew-t variational Bayes filter.

However, the authors of this paper assume the measurement function to be linear. The filter is applied for positioning in an UWB network in [67], using time-of-arrival based range measurements. In order to enable the use of the filter from [66], in [67] a linearised version of the measurement function, linearised at the prior mean position estimate, is used. Using a linearised measurement function might result in information loss.

7 Appendix: Probability Distributions

Normal distribution

A random variable $\underline{z}$ is said to have a normal distribution with mean μ and variance σ^2 if its PDF is of the form [14, p. 232]

$$p_{\underline{z}}(z) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(z-\mu)^2}{2\sigma^2}}. \tag{49}$$

The normal distribution is denoted $\underline{z} \sim \mathrm{N}(\mu, \sigma^2)$.

Truncated normal distribution

A random variable $\underline{z} \sim \mathrm{N}(\mu, \sigma^2)$ limited to $[a, b]$ is said to have a truncated normal distribution with its PDF being

$$p_{\underline{z}}(z) = \frac{\frac{1}{\sigma}\phi\left(\frac{z-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} \tag{50}$$

for $a \leq z \leq b$ and $p_{\underline{z}}(z) = 0$ otherwise [42, p. 156]. ϕ denotes the PDF and Φ the cumulative distribution function (CDF) of $\underline{z} \sim \mathrm{N}(0, 1)$.

Multivariate normal distribution

A $n_{\mathbf{z}}$-dimensional random vector $\underline{\mathbf{z}}$ is said to have a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ if its PDF is of the form [14, p. 236]

$$p_{\underline{\mathbf{z}}}(\mathbf{z}) = \frac{1}{(2\pi)^{n_{\mathbf{z}}/2}|\boldsymbol{\Sigma}|^{1/2}} \mathrm{e}^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})}, \tag{51}$$

where $|\boldsymbol{\Sigma}|$ is the determinant of $\boldsymbol{\Sigma}$. The multivariate normal distribution is denoted $\underline{\mathbf{z}} \sim \mathrm{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Gaussian mixture

A Gaussian mixture is a convex combination of N, possibly multivariate, normal density functions $\mathrm{MVN}(\mathbf{z}; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$, i.e.

$$p_{\underline{z}}(z) = \sum_{j=1}^{N} \omega_j \mathrm{MVN}(\mathbf{z}; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j), \tag{52}$$

where weights ω_j are nonnegative and sum to one. A GM is a valid density function [9, 79].

Gamma distribution

A random variable $\underline{z}$ is said to have a Gamma distribution with shape s and scale θ if its PDF is of the form [83, p. 195]

$$p_{\underline{z}}(z) = \frac{z^{s-1}}{\Gamma(s)\theta^s} \mathrm{e}^{\frac{-z}{\theta}} \tag{53}$$

for $z \in (0, \infty)$, where Γ is the gamma function. Both shape and scale have to be positive. The Gamma distribution is denoted $\underline{z} \sim \mathrm{Gamma}(s, \theta)$.

Uniform distribution

A random variable $\underline{z}$ is said to have a continuous uniform distribution on $[a, b]$ if its PDF is of the form [83, p. 171]

$$p_{\underline{z}}(z) = \frac{1}{b-a} \tag{54}$$

for $a \leq z \leq b$ and zero otherwise. The uniform distribution is denoted $\underline{z} \sim \mathrm{U}(a, b)$.

t-distribution

A random variable $\underline{z}$ is said to have a t-distribution with location ξ, scale σ^2, and ν degrees-of-freedom (DOF) if its PDF is of the form [83, p. 258]

$$p_{\underline{z}}(z) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\sigma\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{1}{\nu}\left(\frac{z-\xi}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}} \tag{55}$$

The t-distribution is denoted $\underline{z} \sim \mathrm{T}(\xi, \sigma^2, \nu)$.

For $\nu \to \infty$ the t-distribution reduces to a normal distribution with ξ and variance σ^2, i.e. $\mathrm{N}(\xi, \sigma^2)$. Distribution $\mathrm{T}(0, 1, \nu)$ is called the standardised t-distribution.

Skew-t distribution

A random variable $\underline{z}$ is said to have a skew-t distribution with location ξ, scale σ^2, skewness λ and ν DOF if its PDF is of the form [54]

$$p_{\underline{z}}(z) = \frac{2}{\sigma} t_\nu\left(\frac{z-\xi}{\sigma}\right) T_{\nu+1}\left(\lambda \frac{z-\xi}{\sigma} \sqrt{\frac{\nu+1}{\nu + \frac{(z-\xi)^2}{\sigma^2}}}\right), \tag{56}$$

where t_ν and $T_{\nu+1}$ denote the PDF and the CDF of the standardised t-distribution. The skew-t distribution is denoted $\underline{z} \sim \mathrm{ST}\left(\xi, \sigma^2, \lambda, \nu\right)$.

Extensive discussions of the skew-t distribution can be found, for example, in [14, p. 101 ff.] and [54]. Here only some important properties are presented.

For $\lambda = 0$ the skew-t distribution reduces to the t-distribution $\mathrm{T}(\xi, \sigma^2, \nu)$. $\mathrm{ST}(0, 1, \lambda, \nu)$ is called a standardised skew-t distribution.

In [P4] and [R1] samples have to be drawn from skew-t distributions. An effective way of sampling from $\mathrm{ST}\left(\xi, \sigma^2, \lambda, \nu\right)$ is to use a hierarchical model of the skew-t distribution. Let $\underline{\tau} \sim \mathrm{Gamma}(\nu/2, \nu/2)$ and $\underline{w} \sim \mathrm{N}(0, 1)$. Then $\underline{t}|(\underline{\tau} = \tau) = |\sigma\, \underline{w}/\sqrt{\tau}|$ is a half-normal random variable with PDF

$$p_{\underline{t}}(t) = 2\frac{\sqrt{\tau}}{\sigma}\phi\left(t\frac{\sqrt{\tau}}{\sigma}\right)[t > 0], \tag{57}$$

where $[\,\cdot\,]$ is the Iverson bracket and ϕ denotes the PDF of the standard normal distribution. Therefore, samples from $\mathrm{ST}\left(\xi, \sigma^2, \lambda, \nu\right)$ can be drawn from the conditional distribution

$$\underline{z}|(\underline{t} = t, \underline{\tau} = \tau) \sim \mathrm{N}\left(\xi + \frac{\lambda t}{\sqrt{1+\lambda^2}}, \frac{1-\delta_\lambda^2}{\tau}\sigma^2\right). \tag{58}$$

References

[1] JAGS - just another Gibbs sampler, September 2013. URL `http://mcmc-jags.sourceforge.net`.

[2] 3GPP TS 36.305. E-UTRAN; Stage 2 functional specification of UE positioning in E-UTRAN. Technical report, Rel. 12, 2014.

[3] B. Alavi and K. Pahlavan. Modeling of the TOA-based distance measurement error using UWB indoor radio measurements. *IEEE Communications Letters*, 10(4):275–277, April 2006.

[4] S. Ali-Löytty. Efficient Gaussian mixture filter for hybrid positioning. In *2008 IEEE/ION Position Location and Navigation Symposium*, pages 60–66, Monterey, California, May 2008.

[5] S. Ali-Löytty. *Gaussian Mixture Filters in Hybrid Positioning*. PhD thesis, Tampere University of Technology, August 2009. URL `http://URN.fi/URN:NBN:fi:tty-200905191055`.

[6] S. Ali-Löytty and N. Sirola. Gaussian mixture filter and hybrid positioning. In *Proceedings of ION GNSS 2007*, pages 562–570, Fort Worth, Texas, September 2007.

[7] S. Ali-Löytty and N. Sirola. Gaussian mixture filter in hybrid navigation. In *Proceedings of The European Navigation Conference GNSS 2007*, pages 831–837. Geneva, Switzerland, May 2007.

[8] S. Ali-Löytty, N. Sirola, and R. Piché. Consistency of three Kalman filter extensions in hybrid navigation. In *Proceedings of The European Navigation Conference GNSS 2005*. Munich, Germany, July 2005.

[9] D. L. Alspach and H. W. Sorenson. Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE Transactions on Automatic Control*, 17(4):439–448, August 1972.

[10] I. Arasaratnam and S. Haykin. Square-root quadrature Kalman filtering. *IEEE Transactions on Signal Processing*, 56(6):2589–2593, June 2008.

[11] I. Arasaratnam and S. Haykin. Cubature Kalman filters. *IEEE Transactions on Automatic Control*, 54(6):1254–1269, June 2009.

[12] I. Arasaratnam, S. Haykin, and R. J. Elliott. Discrete-time nonlinear filtering algorithms using Gauss–Hermite quadrature. *Proceedings of the IEEE*, 95(5):953 – 977, May 2007.

[13] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50 (2):174–188, 2002.

[14] A. Azzalini. *The skew-normal and related families*. Cambridge University Press, 2014.

[15] Y. Bar-Shalom, R. X. Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation, Theory Algorithms and Software*. John Wiley & Sons, 2001.

[16] P. Bellavista, A. Küpper, and S. Helal. Location-based services: Back to the future. *IEEE Pervasive Computing*, 7(2):85–89, 2008.

[17] S. M. Bo-Chieh Liu, K.-H. Lin, and J.-C. Wu. Analysis of hyperbolic and circular positioning algorithms using stationary signal-strength-difference measurements in wireless communications. *IEEE Transactions on Vehicular Technology*, 55(2): 499–509, March 2006.

[18] M. A. Caceres, F. Sottile, and M. A. Spiritio. Adaptive location tracking by kalman filter in wireless sensor networks. In *2009 IEEE International Conference on Wireless and Mobile Computing, Networking and Communications*, pages 123–128. Marrakech, October 2009.

[19] J. M. Castro-Arvizu, J. Vilà-Valls, P. Closas, and J. A. Fernández-Rubio. Simultaneous tracking and RSS model calibration by robust filtering. In *2014 48th Asilomar Conference on Signals, Systems and Computers*, pages 706–710. Pacific Grove, CA, November 2014.

[20] J. Chung, M. Donahoe, C. Schmandt, I.-J. Kim, pedram Razavai, and M. Wiseman. Indoor location sensing using geo-magnetism. In *Proc. 9th International Conference on Mobile Systems, Applications and Services*, June-July 2011.

[21] P. Closas, C. Fernández-Prades, and J. Vilà-Valls. Multiple quadrature Kalman filtering. *IEEE Transactions on Signal Processing*, 60(12):6125–6137, December 2012.

[22] I. Constandache, S. Gaonkar, M. Sayler, R. R. Choudhury, and L. Cox. EnLoc: Energy-efficient localization for mobile phones. In *2009 Proceedings IEEE INFOCOM*, pages 2716–2720. Rio de Janeiro, Brazil, April 2009.

[23] D. F. Crouse, P. Willett, K. Pattipati, and L. Svensson. A look at Gaussian mixture reduction algorithms. In *Proceedings of the 14th International Conference on Information Fusion (FUSION '11)*. Chicago, IL, USA, 2011.

[24] D. Dardari, A. Conti, U. Ferner, A. Giorgetti, and M. Z. Win. Ranging with ultrawide bandwidth signals in multipath environments. *Proceedings of the IEEE*, 97(2):404–426, February 2009.

[25] D. Dardari, P. Closas, and P. M. Djurić. Indoor tracking: Theory, methods, and technologies. *IEEE Transactions on Vehicular Technology*, 64(4):1263–1278, April 2015.

[26] P. Davidson. *Algorithms for Autonomous Personal Navigation Systems*. PhD thesis, Tampere University of Technology, November 2013. URL `http://urn.fi/URN:ISBN:978-952-15-3215-3`.

[27] J. A. del Peral-Rosado, J. A. López-Salcedo, G. Seco-Granados, F. Zanier, and M. Crisci. Analysis of positioning capabilities of

3GPP LTE. In *Proceeddings of 25th International Technical Meeting of the Satellite Division of ION*, Nashville, TN, USA, September 2012.

[28] B. Dil and P. Havinga. On the calibration and performance of RSS-based localization methods. In *Internet of Things 2010 Conference*. Tokyo, Japan, November/December 2010.

[29] S. Frühwirth-Schnatter and S. Pyne. Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions. *Biostatistics*, 11(2):317–336, April 2010. doi:10.1093/biostatistics/kxp062.

[30] C. Gao and R. Harle. Sequence-based magnetic loop closures for automated signal surveying. In *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. Banff, AB, October 2015.

[31] A. E. Gelfand and A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409, 1990.

[32] S. Gezici, Z. Tian, G. B. Biannakis, H. Kobayashi, A. F. Molisch, H. V. Poor, and Z. Sahinoglu. Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks. *IEEE Signal Processing Magazine*, 22(4):70–84, July 2005.

[33] H. L. Groginsky. Position estimation using only multiple simultaneous range measurements. *IRE Transactions on Aeronautical and Navigational Electronics*, ANE-6(3):178–187, September 1959.

[34] F. Gustafsson and F. Gunnarsson. Mobile positioning using wireless networks: possibilities and fundamental limitations based on available wireless network measurements. *IEEE Signal Processing Magazine*, 22(4):41–53, July 2005.

[35] S. Han, Z. Gong, W. Meng, and C. Li. An indoor radio propagation model considering angles for WLAN infrastructures. *Wireless Communications and Mobile Computing*, 15(16):2038–2048, November 2015.

[36] U. D. Hanebeck and J. Steinbring. Progressive Gaussian filtering based on Dirac mixture approximations. In *Proceedings of the 15th International Conference on Information Fusion (Fusion 2012)*, pages 1697–1704. Singapore, July 2012.

[37] R. Harle. A survey of indoor inertial positioning systems for pedestrians. *IEEE Communication Surveys and Tutorials*, 15(3): 1281–1293, July-September 2013.

[38] V. Honkavirta, T. Perälä, S. Ali-Löytty, and R. Piché. A comparative survey of WLAN location fingerprinting methods. In *Proceedings of the 6th Workshop on Positioning, Navigation and Communication 2009 (WPNC'09)*, pages 243–251. Hannover, Germany, March 2009.

[39] K. Ito and K. Xiong. Gaussian filters for nonlinear filtering problems. *IEEE Transactions on Automatic Control*, 45(5):910–927, May 2000.

[40] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*, volume 64 of *Mathematics in Science and Engineering*. Academic Press, 1970.

[41] B. Jia, M. Xin, and Y. Cheng. Sparse-grid quadrature nonlinear filtering. *Automatica*, 48(2):327–341, February 2012.

[42] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous univariate distributions*, volume 1. John Wiley & Sons, 2nd edition, 1994.

[43] D. B. Jourdan, D. Dardari, and M. Z. Win. Position error bound for UWB localization in dense cluttered environments. *IEEE Transactions on Aerospace and Electronic Systems*, 44(2):613–628, April 2008.

[44] S. J. Julier and J. K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, March 2004.

[45] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte. A new approach for filtering nonlinear systems. In *American Control Conference*, volume 3, pages 1628–1632, 1995.

[46] R. Kaneto, Y. Nakashima, and N. Babaguchi. Real-time user position estimation in indoor environments using digitalwatermarking for audio signals. In *2010 20th International Conference on Pattern Recognition (ICPR)*. Istanbul, August 2010.

[47] M. Kok, J. D. Hol, and T. B. Schön. Indoor positioning using ultrawideband and inertial measurements. *IEEE Transactions on Vehicular Technology*, 64(4):1293–1303, April 2015.

[48] L. Koski. Positioning with Bayesian coverage area estimates and location fingerprints. Master's thesis, Tampere University of Technology, http://math.tut.fi/posgroup/koski_mscth.pdf, March 2010.

[49] L. Koski, T. Perälä, and R. Piché. Indoor positioning using WLAN coverage area estimates. In *2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. Zurich, Switzerland, September 2010.

[50] L. Koski, R. Piché, V. Kaseva, S. Ali-Löytty, and M. Hännikäinen. Positioning with coverage area estimates generated from location fingerprints. In *Proceedings of the 7th Workshop on Positioning, Navigation and Communication 2010 (WPNC'10)*, pages 99–106. Dresden, Germany, March 2010.

[51] M. Kristan, D. Skočaj, and A. Leonardis. Incremental learning with Gaussian mixture models. In *Computer Vision Winter Workshop 2008*, pages 25–32. Moravske Toplice, Slovenia, February 2008.

[52] B. Li, T. Gallagher, A. G. Dempster, and C. Rizos. How feasible is the use of magnetic field alone for indoor positioning? In *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. Sydney, NSW, Australia, November 2012.

[53] X. Li. RSS-based location estimation with unknown pathloss model. *IEEE Transactions on Wireless Communications*, 5(12): 3626–3633, December 2006.

[54] T. I. Lin, J. C. Lee, and W. J. Hsieh. Robust mixture modeling using the skew t distribution. *Statistics and Computing*, 17(2): 81–92, June 2007.

[55] H. Liu, H. Darabi, P. Banerjee, and J. Liu. Survey of wireless indoor positioning techniques and systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 37(6):1067–1080, November 2007.

[56] X. Liu, H. Makino, and Y. Maeda. Basic study on indoor location estimation using visible light communication platform. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2377–2380. Vancouver, BC, August 2008.

[57] J. T.-H. Lo. Finite-dimensional sensor orbits and optimal nonlinear filtering. *IEEE Transactions on Information Theory*, 18(5): 583–588, September 1972.

[58] S. M. Lynch, editor. *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. Springer New York, 2007.

[59] R. Mautz. Indoor positioning technologies. Habilitation Thesis, February 2012.

[60] G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*, volume 2. John Wiley & Sons, 2008.

[61] A. F. Molisch. *Wireless Communications*. Wiley - IEEE, 2 edition, January 2011.

[62] P. Müller, M. Raitoharju, and R. Piché. A field test of parametric WLAN-fingerprint-positioning methods. In *Proceedings of the 17th International Conference on Information Fusion (Fusion 2014)*, pages 1 – 8. Salamanca, Spain, July 2014.

[63] A. Mulloni, D. Wagner, I. Barakonyi, and D. Schmalsteig. Indoor positioning and navigation with camera phones. *IEEE Pervasive Computing*, 8(2):22–31, April–June 2009.

[64] K. P. Murphy. *Machine Learning - A Probabilistic Perspective*. The MIT Press, 2012.

[65] H. Nurminen, J. Talvitie, S. Ali-Löytty, P. Müller, E. S. Lohan, R. Piché, and M. Renfors. Statistical path loss parameter estimation and positioning using RSS measurements in indoor

wireless networks. In *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. Sydney, NSW, Australia, November 2012.

[66] H. Nurminen, T. Ardeshiri, R. Piché, and F. Gustafsson. Robust inference for state-space models with skewed measurement noise. *IEEE Signal Processing Letters*, 22(11):1898–1902, November 2015.

[67] H. Nurminen, T. Ardeshiri, R. Piché, and F. Gustafsson. A NLOS-robust TOA positioning filter based on a skew-t measurement noise model. In *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–7. Banff, Alberta, Canada, October 2015.

[68] J. Park, D. Curtis, S. Teller, and J. Ledlie. Implications of device diversity for organic localization. In *2011 Proceedings IEEE INFOCOM*, pages 3182–3190, April 2011.

[69] R. Piché. Estimation of model parameters. In *Mathematical Modeling with Multidisciplinary Applications*, pages 169–190. John Wiley and Sons, 2013.

[70] R. Piché, S. Särkkä, and J. Hartikainen. Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate student-t distribution. In *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6, 2012.

[71] C. P. Robert and G. Casella. *Monte Carlo statistical Methods*, volume 3. Springer, 1999.

[72] T. Roos, P. Myllymäki, and H. Tirri. A statistical modeling approach to location estimation. *IEEE Transactions on Mobile Computing*, 1(1):59–69, Jan-Mar 2002.

[73] D. J. Salmond. Mixture reduction algorithms for target tracking. In *IEE Colloquium on State Estimation in Aerospace and Tracking Applications*, pages 7/1–7/4. London, England, 1989.

[74] S. Särkkä. *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.

[75] D. Schieferdecker and M. F. Huber. Gaussian mixture reduction via clustering. In *Proceedings of the 12th International Conference on Information Fusion (FUSION '09)*, pages 1536–1543. Seattle, WA, USA, 2009.

[76] S. Y. Seidel and T. S. Rappaport. 914 MHz path loss prediction models for indoor wireless communications in multifloored buildings. *IEEE Transactions on Antennas and Propagation*, 40 (2):207–217, February 1992.

[77] S. Shrestha, J. Talvitie, and E. S. Lohan. Deconvolution-based indoor localization with WLAN signals and unknown access point locations. In *2013 International Conference on Localization and GNSS (ICL-GNSS)*, pages 1–6. Torino, Italy, June 2013.

[78] M. I. Silventoinen and T. Rantalainen. Mobile station emergency locating in GSM. In *Proceedings of IEEE International Conference on Personal Wireless Communications*, pages 232–238. New Delhi, India, February 1996.

[79] H. W. Sorenson and D. L. Alspach. Recursive Bayesian estimation using Gaussian sums. *Automatica*, 7(4):465–479, July 1971.

[80] J. Talvitie, M. Renfors, and E. S. Lohan. Distance-based interpolation and extrapolation methods for RSS-based localization with indoor wireless signals. *IEEE Transactions on Vehicular Technology*, 64(4):1340–1353, April 2015.

[81] E. Trevisani and A. Vitaletti. Cell-id location technique, limits and benefits: an experimental study. In *Proceedings of Sixth IEEE Workshop on Mobile Computing Systems and Applications (WMCSA 2004)*. Windermere, Cumbria, United Kingdom, December 2004.

[82] J. Vilà-Valls, P. Closas, C. Fernández-Prades, and J. A. Fernandez-Rubio. Nonlinear Bayesian filtering in the Gaussian scale mixture context. In *Proc. 20th European Signal Processing Conference (EUSIPCO 2012)*, volume 529–533. Bucharest, Romania, August 2012.

[83] R. E. Walpole, R. H. Myers, S. L. Myers, and K. Ye. *Probability & Statistics for engineers & scientists*. Pearson Education, 8th edition, 2007.

[84] E. A. Wan and R. van der Merwe. The unscented Kalman filter for nonlinear estimation. In *The IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000*, pages 153–158, October 2000.

[85] L. Wirola. *Studies on Location Technology Standards Evolution in Wireless Networks*. PhD thesis, Tampere University of Technology, March 2010. URL `http://URN.fi/URN:NBN:fi:tty-201002121065`.

[86] L. Wirola, I. Halivaara, and J. Syrjärinne. Requirements for the next generation standardized location technology protocol for location-based services. *Journal of Global Positioning Systems*, 7(2):91–103, 2008.

[87] L. Wirola, T. A. Laine, and J. Syrjärinne. Mass-market requirements for indoor positioning and indoor navigation. In *2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. Zurich, Switzerland, September 2010.

PUBLICATION 1

Philipp Müller, Matti Raitoharju, Simo Ali-Löytty, Laura Wirola, and Robert Piché: A Survey of Fingerprinting and Parametric Fingerprint-Positioning Methods. In *Gyroscopy and Navigation*, Vol. 7, No. 2, pages 107–127, April 2016.
DOI: 10.1134/S2075108716020061

ISSN 2075-1087, Gyroscopy and Navigation, 2016, Vol. 7, No. 2, pp. 107–127. © Pleiades Publishing, Ltd., 2016.
Published in Giroskopiya i Navigatsiya, 2016, No. 1, pp. 3–35.

A Survey of Parametric Fingerprint-Positioning Methods[1]

Ph. Müller[a], M. Raitoharju[a], S. Ali-Löytty[b], L. Wirola[a], and R. Piché[a]

[a] *Department of Automation Science and Engineering (ASE), Tampere University of Technology, Tampere, Finland*
[b] *Department of Mathematics, Tampere University of Technology, Tampere, Finland*
e-mail: philipp.muller@tut.fi, matti.raitoharju@tut.fi, robert.piche@tut.fi, simo.ali-loytty@tut.fi

Received August 31, 2015

Abstract—The term fingerprint-based (FP) positioning includes a wide variety of methods for determining a receiver's position using a database of radio signal strength measurements that were collected earlier at known locations. Nonparametric methods such as the weighted k-nearest neighbor (WKNN) method are infeasible for large-scale mobile device services because of the large data storage and transmission requirements. In this work we present an overview of parametric FP methods that use model-based representations of the survey data. We look at three different groups of parametric methods: methods that use coverage areas, methods that use path loss models, and methods that use Gaussian mixtures. Within each group we study different approaches and discuss their pros and cons. Furthermore, we test the positioning performance of several of the analyzed approaches in different scenarios using real-world WLAN indoor data and compare the results to those of the WKNN method.

DOI: 10.1134/S2075108716020061

I. INTRODUCTION

Over the last decades positioning techniques have received extensive attention, because they have become the backbones of an increasing number of position-aware applications in commercial, public service, and military networks [1, 2]. Those applications include vehicle navigation, intelligent transport systems, inventory tracking, vehicle tracking, fleet management, resource management, environment monitoring, emergency services (E911 in North America, E112 in Europe), medical services (e.g. patient and equipment surveillance in hospitals), and rescue operations (e.g. locating fire fighters in burning buildings). For smart phone users applications include location identification, local search, suggesting local points of interest, geo-tagging of photos and videos, location sensitive billing, and targeted advertising [2–8]. Because many of these applications have to run on small mobile devices, the positioning algorithms have strict limits on allowed energy, memory, bandwidth, and computational resources.

In outdoor environments positioning techniques mainly rely on Global Navigation Satellite System (GNSS) signals. Because nowadays almost all new smart phones are equipped with a GNSS receiver precise positioning of the mobile user equipment (UE) outdoors is continuously possible. However, GNSS receivers use significant amounts of energy. Furthermore, indoors and also under forest canopies and in certain urban settings, such as urban canyons, poor signal penetration by GNSS generally results in unavailable or unreliable location information. Therefore, positioning in those environments must rely on other measurements, e.g. from an inertial measurement unit (IMU) or from radio signals such as cellular networks, Bluetooth, wireless local area networks (WLAN), or ultra-wideband (UWB).

Depending on the used algorithm, positioning using cellular telephone measurements can be simple, economic and can be implemented without upgrading UEs or network equipment [3, 9, 10]. Although the cellular network was not originally designed with positioning in mind,[2] it provides in most environments sufficient accuracy (around 100 m indoors and in urban areas possible, around 200 m in suburban areas) for applications such as local search or weather forecast [12]. Since WLANs are ubiquitous in urban areas and the coverage areas of WLAN access points (APs) are much smaller than cellular network cells, WLAN-based positioning techniques have come to be preferred over cellular as alternatives to GNSS [13] when higher accuracy positioning is desired. Like cellular networks, WLANs were not originally designed for positioning. Also WLAN-based positioning uses the already existing infrastructure, and WLAN APs and receivers in UEs are widely available. In the remainder we focus on WLANs although many of the concepts apply also to cellular networks.

[1] The article is published in the original.

[2] An exception is the Long Term Evolution (LTE) standard, which specifies in release 9 the positioning reference signal (PRS), which makes a positioning accuracy on the order of 10 m possible [11].

Most WLAN-based positioning algorithms exploit the correlation between the received signal strength (RSS) and the UE's location (see e.g. [14, p. 47]). Because modeling signal propagation, especially in indoor environments, is rather complex, nonparametric location fingerprinting methods are widely applied for positioning [15]. Those methods estimate the user's position by comparing the list of current AP received signal strength indicator (RSSI) or RSS measurements to a database (called a radio map) of information (called fingerprints) on APs and their corresponding signal strength values for known positions.

Parametric methods include various approaches. In contrast to the nonparametric methods, they only store some parameters (e.g. the parameters of a signal propagation model) in the radio map that summarize the fingerprints (also known as allocation reports, reception reports, or observations) in a certain way, reducing the radio map's size significantly. The user's position can be estimated using parametric methods, for example, by computing distance estimates between the user and the APs using the received RSS/RSSI measurements in a signal propagation model. Within this paper we present an overview of these parametric methods, and compare them with each other and a widely-used nonparametric method.

The main contribution of our paper is to give an overview about recent developments in the field of parametric fingerprint-positioning methods, and analyses their strengths and weaknesses. For nonparametric fingerprinting methods overviews can be found, for example, in [15, 16].

The results of the field tests have been already published in [17]. However, in this paper we provide a more detailed analysis of the results and the parametric methods that were compared. Furthermore, we analyze additional parametric methods, which have not been considered in [17], and compare them with those considered in [17].

The outline of this paper is as follow. We discuss similarities and differences of nonparametric and parametric fingerprinting methods as well as methods used for fingerprint collection and related issues in Section II. The parametric methods presented use Bayes' rule and Bayesian filtering, thus we briefly summarize the idea of positioning using Bayes' rule and Bayesian filtering in Section III. In Section IV we look at parametric FP approaches that use rather simple models for describing the area that an AP covers, i.e. in which it can be heard. Those methods reduce the radio map's size tremendously while providing sufficient accuracy for many applications in the positioning phase. Section V is dedicated to signal propagation path loss models, which are calibrated from FP data. In Section VI we present an overview of approaches that rely (partly) on parametric fingerprinting techniques, and use mixtures of distributions for modeling position estimates etc. These techniques are useful for nonlinear and/or non-Gaussian systems for which traditional approaches such as Kalman filter (KF) and extended KF (EKF) perform poorly. The performance of a selection of different parametric FP techniques, with and without filtering, is compared in Section VII using benchmark tests using real-world WLAN measurements in indoor environments. Section VIII summarizes and concludes.

Notation: Scalar variables are italic, $\mathbf{x}$ denotes column vectors, and $\mathbf{H}$ denotes a matrix.

LIST OF USED ABBREVIATIONS

AP—access point
AP-ID—access point identifier
AS—adaptive splitting
CA—coverage area
CN—communication node
EGMF—efficient Gaussian mixture filter
EKF—extended Kalman filter
EM—expectation maximisation
FP—fingerprint
GGM—generalised Gaussian mixture
GGMF—generalised Gaussian mixture filter
GM—Gaussian mixture
GMA—Gaussian mixture allowing negative weight
GMB-REM—Gaussian mixture Bayes' with regularised expectation maximisation
GMEM—signal strength estimation model from [44]
GMF—Gaussian mixture filter
GNSS—global navigation satellite system
ID—identifier
IMU—inertial measurement unit
IRLS—iterative reweighted least squares
KF—Kalman filter
KNN—k-nearest neighbour
LS—least squares
MAC—media access point
ML—maximum likelihood
MMSE—minimum mean square error
NN—nearest neighbour
non-LOS—non-line of sight
PL—path loss
RSS—received signal strength
RSSI—received signal strength indicator
SPGMF—sigma point Gaussian mixture filter
SSM—state space model
TP—test point
UE—user equipment
UKF—unscented Kalman filter
UWB—ultra-wideband
WKNN—weighted k-nearest neighbour
WLAN—wireless local area network

II. WHAT ARE PARAMETRIC AND NONPARAMETRIC APPROACHES?

In this section we define what we mean by nonparametric and parametric fingerprinting positioning methods, what they have in common and how they differ. Furthermore, we will discuss issues and possible solutions related to fingerprinting.

The aim of both parametric and nonparametric positioning methods is to determine the $n_{\mathbf{x}}$-dimensional state $\mathbf{x}_k \in \mathbb{R}^{n_{\mathbf{x}}}$ given the $n_{\mathbf{y}}$-dimensional measurements in vector $\mathbf{y}_k \in \mathbb{R}^{n_{\mathbf{y}}}$. The fingerprinting positioning methods have an offline and an online phase. In the offline phase the state $\mathbf{x}_k$ denotes a vector of parameters, while in the online phase the state $\mathbf{x}_k$ denotes the UE position at time t_k and possibly additional information such as the UE velocity.

The measurements in the offline are so-called fingerprints (FPs) collected at known locations. Therefore, they are sometimes also called location FPs. For WLAN-based indoor localization, FPs are generally collected in grid points with one grid point per square meter [18]. The FP data for one of these points consists of identifiers (IDs) of APs from which signals are received by the UE in that specific grid point together with the corresponding signal strength values. From the collected FPs a radio map is generated. The radio map generation of various parametric and nonparametric methods is explained in the remainder of this paper.

In the online phase, measurements in form of a FP are collected in the UE's unknown location. The FP's data depends on the UE location. This dependency between UE position $\mathbf{x}_k$ and measurements $\mathbf{y}_k$ is used for estimating the UE position, and the estimation process uses the radio map entries. In the remainder of this paper several methods for determining position estimates are explained and analyzed.

A. Parametric vs. Nonparametric Fingerprinting

Nonparametric fingerprinting methods use radio maps in which FPs $\mathbf{y}_k$ are stored. For a WLAN the radio map contains, in general, location coordinates, IDs of APs from which signals were received in this location and corresponding signal strength values. In the positioning phase the UE's measurement (AP-IDs and corresponding RSS values) are compared with the radio map entries to infer a position estimate. The simplest approach is the nearest neighbor (NN) method. It returns the location of the FP from the radio map whose measurement is most similar to the UE's measurement as position estimate. This FP is called the nearest neighbor and is found by optimizing a given cost function [19].

A more advanced, widely applied version of the NN is the weighted k-nearest neighbor (WKNN) method. Here, the position estimate is the weighted mean of k locations whose FPs are most similar to the UE's measurement [15]. According to Liu et al. [7] the WKNN combines medium complexity and medium cost with a good robustness and accuracy. Therefore, we use it in Section VII for comparison with parametric localization methods. For overviews on nonparametric location fingerprinting methods we refer the reader to [15, 20] and references therein.

Parametric fingerprinting methods use radio maps in which parameters $\mathbf{x}_k$ are stored that summarize the FP data $\mathbf{y}_k$. Instead of having one FP per grid point, the radio map contains a set of parameters for each AP observed in the FP data. For positioning various methods, dependent on the parameters used in the radio map, can be used. In the following sections we will discuss those methods further.

Nonparametric fingerprinting methods have the advantage that modeling the signal propagation is not needed. These algorithms have been shown to be reasonably precise and reliable in indoor environments (see e.g. [15, 21]), where non-line of sight (non-LOS) situations are very common. Many parametric fingerprinting methods require modeling the signal propagation, which causes problems in challenging indoor environments. In the latter sections of this paper we will consider this topic in more detail.

A major difference between parametric and nonparametric methods is that the radio map's size depends for nonparametric methods on the number of FPs (i.e. grid points), while for parametric methods it depends on the number of APs and number of parameters stored per AP. This means, if more FP data is collected to improve the radio map quality the map's size increases for nonparametric methods while for parametric methods it stays the same, provided that no additional APs are detected.

To get a better understanding on what kind of sizes we are talking it should be noted that for each grid point around 100 samples have to be collected [18] to obtain a reliably FP for the radio map. Since nonparametric FP positioning works directly with the FP data, the size of this data can be a critical issue when FP-based positioning is offered as a large-scale service for mobile devices. Wirola et al. [12] point out that a radio map of just 0.1% of the earth's surface (approximately 130000 km^2 or the size of Greece or Louisiana, USA), with an average density of one FP per m^2, and at least one 6-byte AP Media Access Point (MAC) address per point, needs at least 780 GB. For indoor environments often signals from more than 5 APs can be received. In addition, signal strength values are stored besides the APs' MAC addresses.

One approach for shrinking the radio map for nonparametric method is the usage of data-compression techniques [22, 23]. A more fundamental way to address the issue is to use parametric (model-based) FP methods. Since an AP's signal is generally receivable in many grid points the radio map for parametric methods is in general significantly smaller [17].

Before moving on to the discussion of various parametric methods let us consider how FPs are collected and some of the issues plus solutions related to it.

B. Fingerprinting and Related Issues

For location fingerprinting the fingerprints are collected in an offline phase by site survey, war-driving or crowd-sourcing. In addition to the UE's current position, each fingerprint contains radio characteristic records. In WLANs those records, in general, contain at least AP-IDs and corresponding RSS or RSSI values. Site survey means that the FPs are collected at various locations for generating a fingerprint database. Most often the location has to be entered by hand. This is a significant difference to war-driving, where the location generally is a GNSS-based estimate. As the name suggests war-driving is the act of collecting measurements in a moving vehicle. A disadvantage of both site survey and war-driving is that the data collecting is tedious and expensive. Thus, crowd-sourcing is preferred. In this form of data collection several persons (for example the users of the FP-based localization method) are collecting the FPs. In practice a combination of those methods can be used.

In outdoor environments FP data can be collected via crowd-sourcing or war-driving, making radio map updating less laborious. However, due to the absence of GNSS signals this technique cannot be used indoors and more complicated alternatives have to be employed. Note that it is, nevertheless, possible to use FPs without GNSS-based position (so called unlocated FPs) but then the position accuracy decreases [24].

An important issue in fingerprint positioning is maintaining the radio map. Since network topography (APs can be added, removed, moved or modified) and radio environment change constantly, constant updates of the FP radio map are required to prevent performance deterioration [15, 25]. Already small environment changes can have tremendous influence on the measured RSS values. If the FPs were, e.g., collected while the building was empty, and positioning is done in the same building while crowded with people, the RSS values in a certain position will differ significantly due to body shadowing [18]. The influence of the user's body on RSS has been analyzed, e.g., by Kaemarungsi and Krishnamurthy [26].

Device heterogeneity is another problem that has to be considered. It describes the fact that received signal strength values measured by different devices at the same location and time can vary significantly [13, 21, 27–29]. The heterogeneity includes, for example, the fact that many devices show only unitless RSSI values rather than RSS values, which are always in dBm. Different chipset providers use different RSSI scales with different limits and granularity, which hinders comparison of RSS measurements from different devices [27, 29–31]. To circumvent the problem of RSS heterogeneity some authors use either rankings of RSS values in a FP [21, 32] or RSS ratios [31] or RSS differences [33] instead of the measured RSS values. In the following we will distinguish between RSS and RSSI, depending on which values were used by the cited authors or are needed for the specific task.

To resolve the RSSI scaling problem, calibration methods within the devices could be used; see [28] for a brief overview. Earlier proposed techniques use manual calibration, meaning that users collect measurements at known locations in order to calibrate their device. These methods generally use least-square fitted linear calibration functions [27, 30]. Park et al. [13] argue that calibrating the signal strength levels is insufficient and that differences in signal strength dispersion between devices is also important, and they describe a kernel estimation method for device calibration. Approaches for automatic calibration, which enable simultaneous calibration and positioning, and eliminate the requirement of taking measurements at known reference locations, are proposed, for example, in [28] and [29]. In [29] Koski et al. suggest to construct a RSSI histogram from values collected over a long period. By assuming that distributions of RSSI values for all devices are the same up to translation and scale, these device-specific parameters can be fitted from the histogram. Laoudias et al. [28] make a more thorough study of the histogram approach. They show that assuming that the true RSS value distribution is the same for all users, the parameters for a linear mapping of RSS between devices can be estimated from the histogram. In contrast to Park et al. [13], they find that the device heterogeneity effect is sufficiently reduced by a linear mapping, and that even a simple origin-shift suffices. They describe a method to continually update the calibration on-line while positioning, and find that positioning with this method is as accurate as with a manual calibration except at the beginning of the learning phase.

III. POSITIONING AND BAYESIAN FILTERING

All of the parametric methods presented in this paper use Bayes' rule, for determining parameters and static positioning. For filtered positioning problems Bayesian filtering is used. Thus, we now summarize the Bayesian approaches for both static and filtered positioning problems.

For the static case the state of the system $\mathbf{x}_k$ is estimated as follows. The measurements in vector $\mathbf{y}_k$ are used to determine the posterior probability density function (pdf) of the state by applying Bayes' rule

$$p(\mathbf{x}_k|\mathbf{y}_k) \propto p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k), \quad (1)$$

where $p(\mathbf{x}_k)$ can be an uninformative prior pdf, i.e. constant.

The filtered positioning problem is formulated as a Bayesian filtering problem for a discrete-time state space model (SSM). Within this section we consider

the general discrete-time SSM, namely the nonlinear non-Gaussian additive noise system

$$\mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}) + \mathbf{w}_{k-1}, \quad (2a)$$

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k, \quad (2b)$$

where the errors $\mathbf{w}_{k-1}$ and $\mathbf{v}_k$ are assumed to be white, mutually independent and independent of the initial state $\mathbf{x}_0$. The possibly nonlinear functions $\mathbf{f}_{k-1}(.)$ and $\mathbf{h}_k(.)$ are assumed to be known. In the following the pdf's of $\mathbf{w}_k$ and $\mathbf{v}_k$ are denoted $p(\mathbf{w}_k)$ and $p(\mathbf{v}_k)$, respectively. The aim of filtering is to find the conditional probability density function (posterior)

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}),$$

where $\mathbf{y}_{1:k} \triangleq (\mathbf{y}_1,\ldots,\mathbf{y}_k)$. The posterior can be determined recursively according to the following relations [34].

Prediction (prior):

$$p(\mathbf{x}_k|\mathbf{y}_{k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})\,d\mathbf{x}_{k-1}, \quad (3)$$

Update (posterior):

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1})}{\int p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1})\,d\mathbf{x}_{k-1}}, \quad (4)$$

where the transition pdf is $p(\mathbf{x}_k|\mathbf{x}_{k-1}) = p_{\mathbf{w}_k}(\mathbf{x}_k - \mathbf{f}_{k-1}(\mathbf{x}_{k-1}))$ and the likelihood is

$$p(\mathbf{y}_k|\mathbf{x}_k) = p_{\mathbf{v}_k}(\mathbf{y}_k - \mathbf{h}_k(\mathbf{x}_k)). \quad (5)$$

The initial condition for the recursion is given by the pdf of the initial state $p(\mathbf{x}_0|\mathbf{y}_{1:0}) = p(\mathbf{x}_0)$. Point estimates can be computed from the posterior distribution, e.g. posterior mean.

In general and in the cases analyzed in this paper, the conditional probability density function cannot be determined analytically. Because of this, there are many approximative methods to compute the posterior mean (see e.g. [35]). Besides the posterior mean these methods generally yield also a posterior covariance matrix. In case the true state is known we can check whether the posterior is consistent with respect to a consistency test, such as the Gaussian consistency test [36, p. 235 ff.]. The idea of those tests is to assess the accuracy of the state estimate's covariance matrix. For example, in the Gaussian consistency test, for a risk level of 5%, the posterior is said to be consistent if the posterior covariance matrix $\mathbf{P}_k$ and the posterior mean $\hat{\mathbf{x}}_k \in \mathbb{R}^2$ fulfill the inequality $(\hat{\mathbf{x}}_k - \mathbf{x}_k)^T \mathbf{P}_k^{-1}(\hat{\mathbf{x}}_k - \mathbf{x}_k) \le \chi_2^2 = 5.9915$, where $\mathbf{x}_k$ is the true state.

IV. COVERAGE AREA MODELS

A. Coverage Area Estimation

A computationally light method for parametric fingerprinting is proposed in [20, 29]. In order to reduce the size of the FP radio map the authors represent the coverage area (CA; aka reception region) of any communication node (CN; called AP in WLANs or base station in cellular radio networks) by an elliptical probability distribution, for which the distribution's parameters and the location estimates are solvable in closed form. The probability distribution represents only the region in which a signal from the AP can be received; other than an implied reception strength threshold, it gives no information about the RSS. Although real CAs are often irregularly shaped, modeling them with simple shapes makes it possible to keep the database for storing CAs compact while providing acceptable accuracy. Thus, it enables fast transmission to a UE [20, 29] and fast computation of the UE's position, since any ellipse can be represented by five parameters: three parameters for the origin-centred ellipse (2-by-2 symmetric positive definite matrix), and two more parameters to specify the location of the ellipse centre [37]. Koski et al. [29] and Piché [37] point out that also other shapes, such as circles (three parameters) or polygons (at least six parameters) for modeling CAs of CNs could be used. Furthermore, the 3GPP TS 23.032 standard supports the use of geometrical shapes such as ellipses, polygons and ellipsoids [4, p. 98].

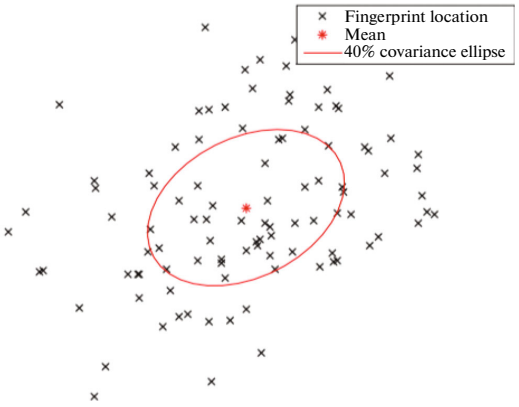


Fig. 1. Fingerprint locations $\mathbf{z}$ in which an AP's signal is received and the mean and covariance of a fitted bivariate Gaussian.

The coverage area is modeled in [20, 29] by a posterior distribution for the ellipse parameters $\boldsymbol{\theta}$ given the FP locations $\mathbf{z} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n\}$ where the CN was heard. The distribution is given by Bayes' rule

$$p(\boldsymbol{\theta}|\mathbf{z}) \propto p(\mathbf{z}|\boldsymbol{\theta})p(\boldsymbol{\theta}), \quad (6)$$

where the likelihood and prior pdf are Gaussian. In other words, the CA is modeled by fitting the mean and covariance of a multivariate Gaussian to the data. Figure 1 shows the FP locations $\mathbf{z}$ as well as the mean and covariance of a fitted multivariate Gaussian, whose parameters are contained in $\boldsymbol{\theta}$ (for the mean and for the symmetric covariance matrix). The ellipse

contains approximately 40% of the bivariate Gaussian's probability.

Using a Bayesian formulation of the regression problem has two advantages. Firstly, the use of the Bayesian prior pdf $p(\boldsymbol{\theta})$ allows one to exploit information about "typical" coverage areas, which is crucial when only a few FPs are available [20, 29]. Such information is available through experimental studies. For example, the typical reception range for WLAN in indoor environments is 20–50 m [6, p. 9]. According to Trevisani and Vitaletti [3] and Molisch [14, p. 17 ff.] the size of a CA in a cellular network depends strongly on the cell type; it may range from 1 m for nanocells to 30 km for macrocells. Furthermore, Trevisani and Vitaletti [3] point out that the size is influenced by a variety of factors, such as interferences, local expected traffic and sensitivity of the UE's antenna. This variability can be also modeled within the Bayesian prior pdf, by using a distribution with a larger variance.

Secondly, using Bayes' rule with the independence assumptions given for the SSM (2) for finding the ellipse's parameters allows recursive estimation and updating of estimates [29]. Updating the posterior $p(\boldsymbol{\theta}|\mathbf{z})$ as new FPs become available can be done by either using Bayes' rule [38, p. 14 ff.] or by computing it as time series [38, p. 29 ff.]. Koski [38] points out that the latter approach enables one to take into account that the parameters can change in time. Such changes are common and can be caused, for example, by constructing new buildings, changing floor plans, or modifying radio network topologies [4, 6, 14, 37, 38].

One possible point of criticism for the method above is the assumption that [20, 29] model location reports having a Gaussian (Normal) distribution. It is common knowledge that the Normal regression model lacks robustness, in the sense that outliers can cause CAs being estimated too large (see e.g. [39]). Reasons for outlier location reports include unusual reception conditions, software or hardware malfunctions in GNSS (when using it for determining the coordinates of the location reports) or radio signal reception [37]. Piché [37] argues that, while gross outliers can easily be detected by heuristics, "moderate" outliers are hard to discover, especially if the list of FPs used for CA determination contains a large amount of them. Thus, he recommends to rather model location reports as having Student-t distribution so that outliers are automatically accommodated by the distribution's heavier tails. In [37] he shows how the Student regression can be computed by Gibbs sampling [40] or Expectation Maximization (EM) algorithm [41]. For both algorithms it is shown in [37] that it is also for Student-t distributed location reports possible to include information on "typical" CAs via an informative prior pdf, as in the case of Gaussian distributed location reports. Introducing such a prior pdf requires only minor changes in EM or Gibbs sampler algorithm.

The methods considered above ignore completely the RSS/RSSI values corresponding to IDs of heard CNs. Hence they are less sensitive to changes in the radio environment than fingerprinting methods that use these values. This gain in robustness, however, comes generally at the cost of lower accuracy compared with nonparametric fingerprinting methods (e.g. WKNN), which besides FP locations and IDs of CNs observed in each FP often also store corresponding RSS or RSSI values. The RSS values, and therefore also the RSSI values, depend on the distance between CN (emitter aka transmitter) and UE (receiver) and are commonly modeled as function of this distance using path loss (PL) models, which will be discussed in section V.

A coverage area method that uses RSS information is proposed in [20, 39]. Instead of storing only one CA per CN in the database, several CAs per CN are stored, which are modeled from FP data that is grouped according to RSS. In [39] the authors examine the use of one, two and three CAs per CN assuming both Gaussian and Student-t distribution for location reports. FPs are grouped based on their RSS values and different CAs are generated using only location reports of their corresponding group. Three different grouping rules are considered: RSS-level, n strongest CNs of each FP and x% strongest CNs of each FP.

B. Positioning Using Coverage Areas

A position estimate for a UE using coverage areas [20, 29, 39] can be obtained by applying Bayes' rule. The position estimate and an uncertainty measure of the estimate can be extracted from a Gaussian posterior probability density function $p(\mathbf{x}|\mathbf{c})$ of the UE position $\mathbf{x}$ given a list $\mathbf{c} = (c_1, c_2, ..., c_N)$ of CNs observed by the UE in the current position. For the conjugate (i.e. Gaussian) prior pdf of this position, a suitable mean and covariance, which represent prior knowledge on UE's position, should be chosen. In case such information is unavailable, setting the covariance very large is justified [20]. For computing the likelihood $p(\mathbf{c}|\mathbf{x})$ [20, 29, 39] it is assumed that prior probabilities of observing c_n are equal for all $n = 1, ..., N$ and that observations are conditionally independent given $\mathbf{x}$.

The latter assumption is a weakness of the model [20]. For neighboring CNs c_i and c_j, for example, the independence assumption is often violated. If c_i is observed in $\mathbf{x}$ then this clearly affects the probability of c_j being observed, since CAs generally overlap, meaning that both CNs can be observed in the same area. Koski et al. [20] speculate that using information on CNs not being observed in $\mathbf{x}$ might improve positioning accuracy significantly. However, because closed-form solutions are unavailable for such data, such an approach could not compete with the proposed algorithm in terms of computational complexity.

Koski et al. [29] point out that WLAN APs (i.e. CNs) are often hearable in an entire building, which results in their coverage areas being useless for positioning. Therefore, eliminating some CNs in both learning and positioning phase may actually increase position accuracy [29], [42, p. 71 f.]. A variety of techniques have been examined for selecting CNs to be pruned, including forward selection and backward elimination, weighting CNs using Generalized Cross Validation, selection of CNs based on information gain, divergence measures or discrimination score. Readers are referred to [43] for a more extensive overview.

In [29] and [44] the authors suggest usage of a signal strength threshold value for reducing the number of observations. CNs heard with an RSSI below this threshold are eliminated from the FP, i.e. are not used for CA determination or positioning. To ensure comparability of RSSI values from different devices the authors in [29] apply the RSSI histogram approach mentioned in Section I. For the positioning using WLAN measurements, positioning mean errors and root mean square errors were reduced significantly by applying this technique. Instead of the histogram approach other calibration methods, which were explained earlier, could be used. However, when using signal strength based elimination, it has to be ensured that CN elimination does not decrease consistency of position estimates and/or overpruning, in which (almost) all CNs are eliminated from the observation list. The latter problem might occur, for example, in outdoor environments for WLAN measurements, where signal strength values and their dynamics are considerably weaker than in indoor environments since the signals generally must pass through thick walls.

This behavior might also explain why using multiple CAs per CN for WLAN-based positioning in outdoor environments provides only small improvements compared with using one CA per CN [39], whereas in indoor environments it yields significantly better accuracy [20, 39]. Alternatively, overpruning might be avoided by using the n strongest rule or the $x\%$ strongest rule, neither of which require RSS calibration.

V. PATH LOSS MODELS

Path loss (PL) models refer to models of the signal power loss L_P or the received signal strength P_{RSS} along a radio link, averaged over large-scale and small-scale fading [14, p. 127]. In the simplest models the PL depends only on the transmit power and the distance d a radio wave travels; more complex models take further factors into account. For an overview of propagation mechanisms and PL models we refer the reader to [6, 14, 45] and references therein.

The relation between the RSS and the radio wave's traveled distance can be used for positioning. From RSS measurements and PL models the distances between a set of reference nodes and the target node are estimated, which then enables estimation of the target node's position. However, the position estimate is sensitive to signal noise and PL model parameter uncertainties because the distance-power gradient is relatively small [1]. Consequently, these estimates are generally less accurate than radio-signal based estimates that are derived using AOA (angle-of-arrival) or time delay measurements. However, Patwari et al. [46] show that for sufficiently high CN-density positioning algorithms relying on PL models (and thus on RSS) can achieve similar performance as time delay based algorithms.

A. Parameter Estimation for PL Models

Earlier studies assumed the parameters of the PL model to be known a-priori, which is an oversimplification for several real-world applications and therefore ill-suited [47]. Thus, the model's parameters should be estimated based on FP data consisting of CN-IDs and corresponding RSS values.

There are various approaches on how to estimate the model parameter(s). Some methods first estimate the CN's position or assume it to be known, and then estimate the parameter(s) using the CN position (e.g. [48, 49]); others estimate the CN's position and the PL model parameter(s) simultaneously (e.g. [47, 50]).

In [48] a statistical modeling approach is introduced, in which the UE's position is estimated using a statistical signal power model, and the CN position is assumed to be known. The parameters of the model are estimated using the EM algorithm [41] to find their maximum likelihood values. The authors point out that the main difference of their algorithm, compared with the geometric approach, is that it infers the signal properties from the location. However, often the positions of CNs are unknown and have to be estimated as well. Li [47] found that estimating the CN location alone for fixed PL model parameters can result in large errors when the values for the parameters are chosen inaccurately, and he recommends simultaneous estimation. Furthermore, he points out that joint estimation removes the necessity of extensive channel measurements.

In [47] Li estimates CN position and the PL exponent n (aka distance-power gradient) of the classic narrowband radio propagation PL model[3]

$$L_P(d) = L_P(1) + 10n\log_{10}(d) + w, \qquad (7)$$

where $L_P(d)$ is the signal power loss in dB at a distance of d meters from the CN. The zero-mean Gaussian random variable w with variance σ_w^2 is used for modeling the shadow fading (aka slow fading). The approxi-

[3] Roos et al. [48] use a more complex PL model that includes a parameter for the transmission direction. Since their focus is on cellular networks in which the CNs (aka base stations), in general, have directional antennas this should provide more accurate PL estimates. For isotropic WLAN APs, as used by Li [47], the PL should (theoretically) be the same in all directions.

mately lognormal distribution of shadow fading, which implies Gaussian distribution of w in (7), has been empirically observed, for example, in [51–54]. For the estimation Li [47] applies the Levenberg-Marquardt method, a modified Gauss-Newton (GN) algorithm, on a system of nonlinear equations.

Nurminen et al. [50] go a step further and estimate in addition the apparent transmission power $A = P_{RSS}(l)$ for a version of the log-distance model, namely

$$P_{RSS}(d) = A - 10n\log_{10}(d) + w, \quad (8)$$

using the Iterative Reweighted Least Squares (IRLS) method, which is also a GN algorithm. According to Dil and Havinga [55] (8) can be used for describing P_{RSS} dependency of distance d in any indoor environment. The Bayesian approach used in [50] furthermore lends itself well to update the estimate of CN position and PL model parameters as new fingerprint data becomes available.

The algorithm uses uninformative Gaussian prior pdf. Given enough fingerprints, according to Nurminen et al. [50], one can choose the valid prior mean values for the PL model parameters arbitrarily, since for large numbers of FPs the posterior distribution is typically unimodal. This is supported by Li's [47] finding that the effect of an inaccurate prior for the PL exponent n on the estimation results is negligibly small. However, he stresses that, especially for cases with limited data, a well-chosen informative prior would be beneficial. Various studies yielded values for the PL exponent for different environments and networks (e.g. [45, 56]); for the apparent transmission power fewer studies are available (e.g. [56]).

For the prior CN position more care should be taken in order to prevent IRLS placing the CN in an area of weak RSS values [50]. Nevertheless, even with such measures it cannot be guaranteed that the algorithms find the correct CN position. For example, the RSS map might contain several peaks or the true CN position could be outside the RSS map or it may be that too few measurements are available for determining a clear peak [50]. Both the (Bayesian) IRLS method in [50] and the Levenberg-Marquardt method in [47] give the user a tool to distinguish between reliable and unreliable position estimates and PL models in the form of a covariance matrix. For the latter approach the covariance can be computed once the optimal estimates are found, in the former it is automatically available as posterior covariance matrix. Furthermore, the approach in [50] accounts for correlation in measurement errors by adding a small constant diagonal matrix for the CN position's covariance matrix (Li [47] assumes all observations to be statistically independent). The cross-correlation between CN position and PL model parameters is, however, neglected, mainly to limit the number of parameters.

One possible point of criticism of the methods in [47, 50] is that the authors assume the standard deviation σ_w of the shadow fading component to be fixed, although [47, 50] stress that it is highly dependent on the propagation environment, which is confirmed, e.g., by Ghassemzadeh et al. for UWB networks [54]. Typical values vary between 1 dB and 6 dB [56, 57] for WLAN. Larger values have been observed, especially in larger buildings (see e.g. [57, p. 139 ff.] for more values). In cellular networks values between 5 dB and 16 dB have been observed [52, 56]. Nurminen et al. [50] use fixed $\sigma_w = 6$ dB, whereas Li [47] studies the influence of varying values that are fixed during the estimation on the errors in CN position estimates. His tests show that the value of σ_w can influence the bias and the efficiency of location estimators significantly depending on the used estimation method.

In [18] Han et al. ignore the shadow fading component's standard deviation, and compute point estimates for the parameters of their PL model. However, their approach is worth consideration because it allows generating a radio map using significantly less FP data. Furthermore, their PL model, which is an extension of Seidel's model [58], takes into account the angles between signal path and obstacles. This means, the modeled PL depends on the angle in which the signal hits the obstacle. For example, if the signal hits a wall in a 90 degree angle the distance it travels through the wall is significantly shorter than if it hits the same wall in a 60 degree angle. The PL model in [18] accounts for this.

For generating the radio map, the authors collect FP data only for a small fraction of the area that should be covered by the radio map. Those FPs are then used to establish a linear equation that can be solved using a least-squares method to obtain the estimates for the PL model parameters. They then use the PL model to generate FPs for those locations in which no measurements were taken. However, this second step will be unnecessary if parametric methods are used for positioning, as will be seen in the next subsection. In their paper Han et al. show that it suffices to collect FPs from 20% of the grid points for which FPs will be available in the radio map. In their test the radio map using those points and their method for generating FPs for the other 80% grid points has a similar cumulative distribution function (cdf) for the predicting error as the radio map for which in all grid points FPs were collected.

Shrestha et al. [19] use deconvolution-based methods to reduce the radio map size by a factor of ten compared with the map for nonparametric approaches. Besides PL model (8), they study also a multi-slope PL model, which takes different values for n depending on the distance between transmitter and receiver. Furthermore, the authors extend both models with an additive floor loss parameter, and consider 3-dimensional positions.

The parameters of the PL models are estimated in [19] as follows. For each FP measurement parameter estimates are computed using either least square (LS), weighted LS or minimum mean square error (MMSE) assuming the FP's location as AP location. Then the

expected RSS values in that location given the current parameter estimates is computed. The FP's location for which the mean square error (MSE) of measured and expected RSS values is smallest is used as AP position estimate. Alternatively, the average of k FPs that give lowest MSE could be chosen. In this case, the PL model parameters have to be recomputed using the new AP position estimate.

B. Positioning Using PL Models

Once the parameters of the PL model and the positions for all CNs are estimated, range estimates can be derived using the PL model and measured RSS values. For computing the UE position estimate subsequently trilateration or some other nonlinear estimation technique can be used.

Nurminen et al. [50] test three different methods that use the PL model (8) with real WLAN data in an indoor office environment: a grid method where Monte Carlo integration is used for computing the likelihood in each point of a spatial grid, the Metropolis-Hastings algorithm, and the IRLS. For comparison the authors apply the CA method presented in [20, 29], filtered using standard Kalman filter, and WKNN with $k = 3$ and unfiltered measurements. Within the tests the floor is assumed to be known. Furthermore, grid method, Metropolis-Hastings algorithm and IRLS are analyzed using both point estimates and Gaussian distributions for the PL model parameter values.

When using extensive FP data for estimation of CN position and parameter estimates, WKNN provides the best accuracy, followed by grid method, Metropolis-Hastings algorithm and IRLS, which all provide similar accuracy. The CA method performs worst. However, when the FP data for some of the CNs is limited, WKNN drops to the same accuracy level as grid method, Metropolis-Hastings algorithm and IRLS. This is in accordance with earlier findings. For example, Dil and Havinga [59] find that in the case of limited FP data nonparametric FP positioning algorithms, such as WKNN, are outperformed by range-based algorithms.

The tests in [50] show that assuming Gaussian distributions for the parameters rather than point estimates is, in general, beneficial; the advantage of using a distribution becomes clearer in the tests with limited FP data. These results do not come as a surprise, since the PL model contains approximation errors [60]. If less FP data is available for estimating the PL model those errors, in general, are larger. Therefore, in such situations it should be beneficial, from a theoretical point of view, to assume more uncertainty in the parameter estimates.

Another possible explanation why assuming a distribution for the PL exponent gives better results than assuming point estimates is that the PL exponent n can be assumed constant only for a limited time in an environment [47]. However, even if the value changes it should still be close to the previous value, as long as the environment stays the same. This can be captured to some extent by assuming some uncertainty in the PL exponent estimate. In addition, storing the uncertainties also enables updating the parameters recursively and using time evolution models when new FPs become available.

In terms of computational demand, the grid method and the Metropolis-Hastings algorithm have no edge compared with the WKNN, whereas the IRLS is significantly faster and achieves running times close to those of the CA method.

As mentioned in the previous subsection, Han et al. [18] use their PL model and some FP data for generating the full radio map of a certain environment. For positioning they then use the (non-parametric) WKNN. Their test shows that positioning accuracy of their approach is slightly worse, but generating the radio map is significantly faster. It could be even faster if the authors would simply store the parameters of their PL model for each CN in their radio map and derive in the positioning phase ranging estimates from the PL model and the RSS values, and then use some nonlinear estimation technique to obtain a UE position estimate.

Shrestha et al. [19] go the same way as Han et al. [18], generating a full radio map from the PL parameters in the positioning phase. Their idea is to create an artificial grid and compute for each grid point signal strength differences between the UE's measured and the expected RSS values in that grid point. The position estimate can then obtained by using a version of the NN method. In their paper the authors use the KNN with $k = 4$. For three of four test buildings the 4NN using the full radio map (i.e. original FP data stored in radio map) outperforms the 4NN using artificial radio map generated from the radio map containing only PL model parameters, but differences are rather small. The best option for computing PL model parameter estimates in the offline-stage is the MMSE, according to the tests. Furthermore, PL model (8) gives the best trade-off between positioning accuracy and computational complexity.

Again, it would be worth to skip the radio map reconstruction step and simply use the PL parameters and measured RSS values to compute range estimates and then apply some nonlinear estimation technique. However, when using the multi-slope PL model using the correct estimate for the PL exponent could be tricky.

Accuracy levels of PL model-based positioning could be improved further by replacing the isotropic PL models that were considered so far with anisotropic PL models. It is well known that channel characteristics in different directions from the CN differ even for omni-directional antennas due to varying environments (e.g., a WLAN AP mounted in corner of a room). Furthermore, in practice directional antennas are widely used since they decrease interference with other systems [14], and allow significantly higher data

throughput and range extension [61]. Thus, using different PL model parameters for different directions might be beneficial in terms of accuracy. However, once more this comes at the cost of a larger database. Storing PL model parameters for two directions already doubles the size of the radio map. In addition, it complicates the positioning phase since one either has to decide which of the parameters to use or has to compute an (weighted) average of the parameters.

VI. GAUSSIAN MIXTURES AND RELATED APPROACHES

A known disadvantage of the CA approach discussed in Section IV is that most of the probability mass is located near the center of the ellipse that is used for describing a CN's coverage area. However, for weak signals the UE is more likely to be close to the edge of the CA. Therefore, CAs yield in such cases rather poor estimates in the positioning phase [20]. In the previous section we looked at approaches that address such issues by taking into account the RSS in addition to the CN-ID by using PL models. Alternatively, we could apply Gaussian mixture (GM) models (aka Gaussian sum models).

A Gaussian mixture is a convex combination of Gaussian density functions $\{\mathrm{N}\}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})$, namely

$$p(\mathbf{x}) = \sum_{n=1}^{N} \omega_n \{\mathrm{N}\}(\mathbf{x};\boldsymbol{\mu}_n,\boldsymbol{\Sigma}_n), \tag{9}$$

where weights ω_n are nonnegative and sum to one. The main motivation behind GM and filters based on it is that any density function can be approximated, except at discontinuities, by a convex combination of Gaussian densities arbitrarily closely [62–64]. That is, as the number of Gaussian components within the GM increases and the norm of all covariance matrices approaches zero the approximative density function converges uniformly towards the desired density function ([62], Lo (1969) and Alspach (1970) cf. [65]). Unlike other approximation techniques such as Gram-Charlier and Edgeworth expansions, a GM is a valid density function itself [62, 65].

Sorenson and Alspach [62] point out that the approximation quality depends not only on the number of components in the GM but also on their placement. Furthermore, they stress that there is no obvious way to choose the parameters of the components due to the GM's lack of orthogonalizability. Thus, they suggest choosing them so that either the L^k-norm of the approximation error is minimized or the GM approximation matches some of the moments of the true density exactly. In addition, they point out that assigning the same covariance to all components eases the computational demand significantly. However, for many cases the latter idea yields poor approximations.

A. Representing FP Data Using GMs

For reducing the size of the FP radio map a more sophisticated method than the single Gaussian coverage area approach presented in [20, 29] that uses signal strength values is proposed by Kaji and Kawaguchi [44]. They suggest representing a CN's RSS distribution as a GM model. Although this approach generally will require more data to be stored in the radio map than the CA approach of Section IV, it should still require considerably less storage compared with traditional FP databases.

In their algorithm the collected FP data is first transformed into a point distribution, where the point density depends on the signal strength received in a FP (the higher the RSS or RSSI value the higher the point density). Then the parameters of the GM model, namely mean values $\{\boldsymbol{\mu}_n\}_{n=1}^N$, covariance matrices $\{\boldsymbol{\Sigma}_n\}_{n=1}^N$ and component weights $\{\omega_n\}_{n=1}^N$, are optimized by EM [41]. Kaji and Kawaguchi point out that their approach allows updating the GM models as new FP data becomes available. They do not provide an equation or rule for determining the number of components N. In our tests in Section VII we use $N = \max([K/40],8)$, where K is the number of FPs in which the specific CN is heard.

A different approach, developed for robot localization, is introduced by Koshizen [66]. The approach is called GM Bayes' with regularized expectation maximization (GMB-REM). In the offline phase of this algorithm fingerprints containing measurements from sensors are collected at various positions within a grid. Then for each position $\mathbf{x}$ in the grid the conditional density function of sensor measurements $\mathbf{y}$ given $\mathbf{x}$ is computed as a regularized GM model. The parameters of the components are chosen such that they maximize the log-likelihood function given the (offline) training data using EM. In the positioning phase the likelihood of each position $\mathbf{x}$ is multiplied with the prior pdf and then renormalized in order to obtain the posterior. In [66–68] the authors introduce further techniques for sensor selection and a new sensor fusion system for the GMB-REM. However, the algorithm suffers from the common drawbacks of FP databases, and its application on large scale is unpractical.

B. Positioning Using GM Models

In positioning tasks the system (2) is often non-Gaussian and/or significantly nonlinear (see e.g. [69] for a criterion for significant nonlinearity). Therefore the Bayesian recursion is generally unsolvable in closed form [62]. Applying an EKF to solve such generally multimodal systems has the disadvantage that it follows a single peak of the pdf, meaning that it gives rather a maximum likelihood estimate than a minimum variance estimator [65]. Besides computationally demanding methods such as the particle filter (see e.g. [70]), also GM-based filters can be used to achieve

excellent performance for significantly nonlinear and/or non-Gaussian systems.

In [65] it is stated that already in 1965 Aoki suggested to approximate the posterior as GM. Ali-Löytty [71] introduces such an approximation that he calls efficient GM filter (EGMF). His method uses parallel planes to split the state space into pieces and then approximates the posterior in every piece by one Gaussian. Ali-Löytty shows that, unlike most other GM filters, the EGMF yields optimal results in the sense of mean and covariance in the linear case. Furthermore, he finds that the EGMF provides better accuracy than traditional Kalman-type filters (e.g. EKF, UKF) and the sigma point GMF (SPGMF) [72], while requiring fewer components than the SPGMF. If the prior pdf follows a Gaussian distribution, then the EGMF's number of components is equal to the dimension of state variable **x**. One disadvantage of the SPGMF is its need for analytical differentiation. In order to avoid such differentiation, Raitoharju and Ali-Löytty [73] propose the adaptive splitting (AS) method. This method first finds the direction of maximal nonlinearity within a Gaussian prior pdf. If the measurement's nonlinearity is significant with respect to the criterion proposed by Ali-Löytty and Sirola [72], then the Gaussian component describing the measurement is split into a mixture of two Gaussians. The splitting is repeated until none of the components shows a significant nonlinearity anymore. The results in [73] suggest that the proposed method requires fewer components than SPGMF while providing better approximation of the reference pdf.

A further important fact is that assuming Gaussian distribution of the prior pdf is not always feasible. It is obvious that, when looking into filtering, for a posterior at time k that is described by a GM, the prior pdf at time $k + 1$ should be also a GM. In general, if the initial state is not Gaussian and/or in highly nonlinear situations we should apply a bank of Gaussian filters, namely a GM filter (GMF), for solving the problem [72]. Anderson and Moore [74, p. 212] and Ali-Löytty [75] show that if the GM approximation of the prior pdf converges towards the true prior pdf as the number of components increases while their covariances decrease, then the GM approximation of the posterior converges towards the true posterior as well. Lo [64] presents an application of GMs for filtering a system with linear dynamics and arbitrarily distributed prior pdf and some examples, which provide an excellent introduction to the concept.

Care has to be taken to limit the number of components in the GM; this has been mentioned already when GM was introduced [62, 65]. Sorenson and Alspach [62] suggest to either merge components with approximately equal means and covariances or drop components with sufficiently small weights on the GM (called forgetting). Alternative methods are introduced, for example, in [69, 76, 77]. For a more detailed and broader overview on component reduction methods we refer the reader to [78, 79] and references therein.

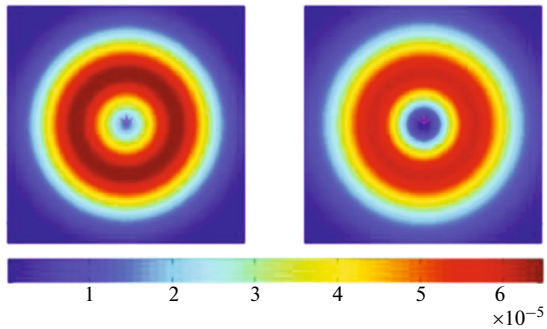


Fig. 2. Normalized exact likelihood (left side) for measurement of an isotropic CN (magenta asterisk inside the ring) and its approximation yielded by GGM (right side).

In addition to reducing the number of Gaussian components it would be beneficial to already keep the number of components small in the approximation phase. In [80] Müller et al. therefore propose a generalized version of GM (GGM) that relaxes the nonnegativity restriction on component weights, and call it Gaussian mixture allowing negative weights (GMA). For the isotropic ranging model, which is considered in [80], the measurement likelihood has a ring-shaped pattern as shown in the left plot of Fig. 2. While a traditional GM would require a large number of components, due to the infinite number of peaks of the likelihood, the GGM yields a satisfying approximation with only two components (see Fig. 2 in the right plot), one having positive and one having negative weight. Special care has to be taken when assigning the weights of the components to ensure that the resulting likelihood is everywhere nonnegative and thus a valid likelihood. In the filter based on GGM the authors collapse the GM posterior at each time step using moment matching, since the reduction methods mentioned previously generally are only suitable for GMs with nonnegative component weights.

Müller et al. apply the GGM in [80] for positioning in cellular networks, and their results indicate that the GGM outperforms both single time-step EKF and the Gaussian CA approach [20, 29] in terms of accuracy and consistency. The filtered version of the GGM (GGMF) also outperforms the EKF and the CA-based filtered approach. In [81]) those findings are confirmed for positioning in an UWB network.

Ali-Löytty and Sirola [69] perform extensive simulations of GM using both cellular measurements and measurements from a Global Navigation Satellite System (GNSS), i.e. hybrid positioning. Their results suggest that only multimodal likelihoods should be approximated by a GM.

Table 1. Data set sizes. Some APs could be heard on several floors and/or in both buildings

Building	Floor	APs	FPs	TPs
1	1	200	889	19
1	2	289	243	47
1	3	212	160	22
2	1	154	1530	168
2	2	186	1582	33
2	3	148	333	19

C. Alternatives to Gaussian Mixtures

All the methods considered in this section so far could be significantly degraded by the previously mentioned sensitivity of Gaussian regression to outliers. Bishop and Svensen [82] point out that this sensitivity might result in an overestimation of the number of required Gaussian components. They propose a Bayesian approach for mixture modeling based on Student-t distributions, which is more robust to non-Gaussianity in the data (McLachlan and Peel [83] make the same proposition). The major drawback of using Student-t distributed components is that, contrary to using Gaussian distributed components, no closed-form solution for the likelihood maximization exists [82, 83]. However, as [37, 82] show, any Student-t distribution can be represented as an infinite mixture of scaled Gaussians. Therefore, EM can be used to find the maximum likelihood, while the computational load of the proposed algorithm [82] is only slightly larger than using the ML technique for finding parameters of GM models. However, t-mixtures have so far not been used in FP positioning.

VII. COMPARATIVE TESTING

In this section we compare the performance of several parametric fingerprinting and positioning methods described in the previous sections. We evaluated these methods by analyzing their WLAN based positioning accuracy for six test tracks located within two buildings of Tampere University of Technology. Building 1 has an area of approximately 10000 m^2 and building 2 has an area of approximately 6600 m^2; both buildings are three-story. The total number of detected APs within both buildings is 506. For two of the tracks measurements were collected several months later than for the other four tracks, which were collected at the same time as the data used for generating the radio maps. Some of the test tracks had floor changes, which were assumed to be known. The radio maps were built separately for each floor. Table 1 shows for each floor of the two buildings the number of detected APs, the number of FPs, and the number of test points (TP) for the four tracks collected at the same time as the data used for the radio maps. TPs are points on the test tracks that we positioned in our evaluation.

For comparison we implemented CA-based positioning with single CA [29] and 2-level CAs with limit −70 dBm [39], PL model [50], GGM approximation of the PL model [80] and the signal strength estimation model from [44] (denoted GMEM). In addition to these parametric methods we used a weighted k-nearest neighbors method (WKNN) with $k = 5$ as a reference. Figure 3 shows how much data

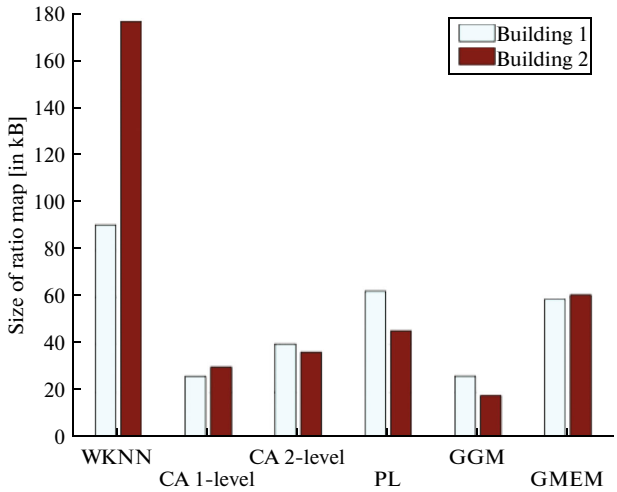


Fig. 3. Data storage requirements for radio maps for tested methods in our two test buildings.

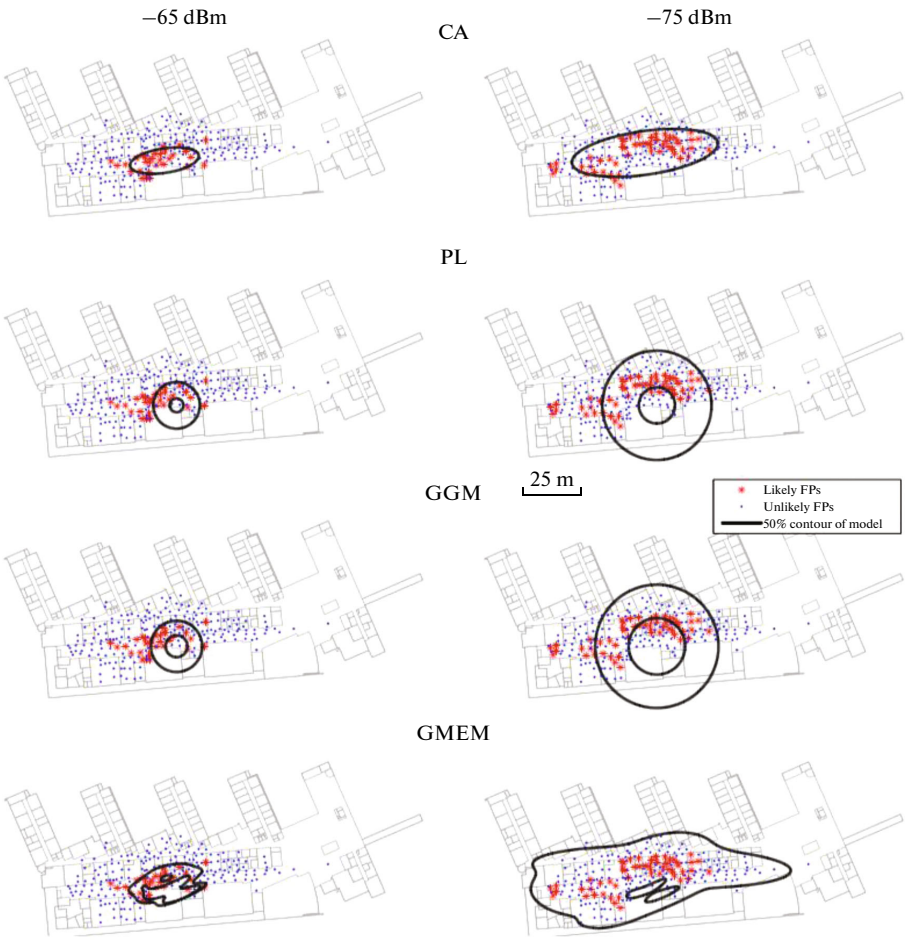


Fig. 4. Likelihoods for different models and different RSS values. Each likely FP has a higher probability mass than any unlikely FP.

storage each method requires for its radio map. The WKNN method does not summarize the FPs in any way and therefore requires the largest radio maps in both buildings. In our tests the parametric methods reduce the radio map size between 30% and 90%. However, because the size of the radio map used by WKNN depends on the number of FPs and the size of the radio map used by the other methods depend on the number of APs these numbers cannot be generalized.

Figure 4 shows the contour curves containing 50% of probability mass for all tested approaches (1-level CA is same as the 2-level model with measurement −75 dBm), except WKNN. For computing the probabilities from likelihoods we used a rectangular uniform prior pdf that covers the whole building. The FPs are similarly divided into groups *likely* and *unlikely* with likely FPs containing 50% of the probability mass. The standard deviation for RSS based methods is set to 6 dB. By visual inspection the shapes of contours are quite different except for PL and GGMF, but if we consider the numbers of likely FPs inside the contours they are similar, except for the GMEM with weak signal strength.

In Figure 5 for all APs the PL exponents estimated by Nurminen's approach [50] as a function of the number of FPs in which the specific AP was observed are displayed. The figure shows that for small numbers of FPs the PL exponents often take values less than 2 (68% of all APs that were received in fewer than 100 FPs have $n < 2$ but only 27% of all APs that were heard in more than 100 FPs have $n < 2$). A PL exponent of 2 means that the signal propagates in free space; values smaller than 2 in our tests can be

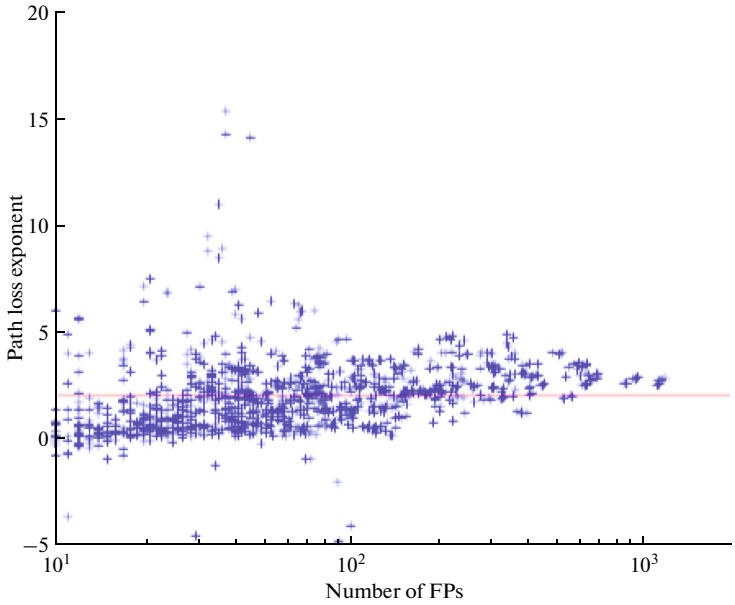


Fig. 5. PL exponent as function of FPs used in learning.

explained by the fact that the corridors in which the FPs were collected acted as waveguides [14, p. 66].

The true routes for all six test tracks were measured by clicking a map plot on a touch screen while walking and interpolating between the clicks, and were estimated for both static case and time series (i.e. filtered case). For the filtering we considered the state vector $\mathbf{x}_k$ containing location and velocity of the UE. Both CA-models and GGMF were updated using a plain Kalman filter. In addition, we collapsed the GGMF to a single component after 5 measurements and after each time step. The GMEM used a grid for static position estimation and a particle filter with 300 particles for the time series estimation; the PL model method used Gauss-Newton for static positioning and a GM filter [73] for time series. In time series the effect of parameter uncertainties varied depending on the location, and therefore was computed in the prior mean of the estimate. The WKNN was given a standard deviation of 10 meters for filtering with a Kalman filter.

For filtering we chose a linear state transition equation (2a) with additive zero-mean noise, i.e. $\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1}$ with

$$\mathbf{F}_{k-1} = \begin{bmatrix} \mathbf{I} & \Delta t\mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \tag{10}$$

where Δt is the measurement interval in seconds, and $\mathbf{w}_{k-1} \sim N(0, \mathbf{Q})$ with

$$\mathbf{Q} = 0.1\mathbf{m}^2 \begin{bmatrix} \frac{(\Delta t)^3}{3}\mathbf{I} & \frac{(\Delta t)^2}{2}\mathbf{I} \\ \frac{(\Delta t)^2}{2}\mathbf{I} & \Delta t\mathbf{I} \end{bmatrix}. \tag{11}$$

The measurement equation (2b) depended on the used positioning approach. For the static WKNN $\mathbf{y}_k$ contained AP-IDs and corresponding RSS values. The k strongest AP-IDs were picked and the weighted average of their locations was used as position estimate. In filtering the weighted average was used as posterior mean for the Kalman filter and the posterior covariance matrix was set to $10^2\mathbf{m}^2\mathbf{I}$.

For the CA approaches, the jth measurement $y_{k,j}$ was modeled as $y_{k,j} = \mathbf{m}_{\mathrm{ID}_{k,j}} + \in$, where $\mathbf{m}_{\mathrm{ID}_{k,j}}$ is the mean of the CA for the AP with identifier $\mathrm{ID}_{k,j}$ and $\in$ is zero-mean Gaussian with the same covariance as the AP's CA model.

In the PL method y_k contained RSS measurements. The RSS for the AP $\mathrm{ID}_{k,j}$ was modeled as

$$y_{k,j} = A_{\mathrm{ID}_{k,j}} - 10n_{\mathrm{ID}_{k,j}} \log_{10}\left(\left\|\mathbf{x}_k - \mathbf{a}_{\mathrm{ID}_{k,j}}\right\|\right) + v_{k,j}, \tag{12}$$

where $\|.\|$ is the Euclidean distance between UE at $\mathbf{x}_k$ and AP $\mathrm{ID}_{k,j}$ located at $\mathbf{a}_{\mathrm{ID}_{k,j}}$. This measurement model was used for both static and filtered positioning.

For the GGM approach the PL model (8) was used to derive the covariance matrices of the Gaussian components. Each measured RSS value was modeled

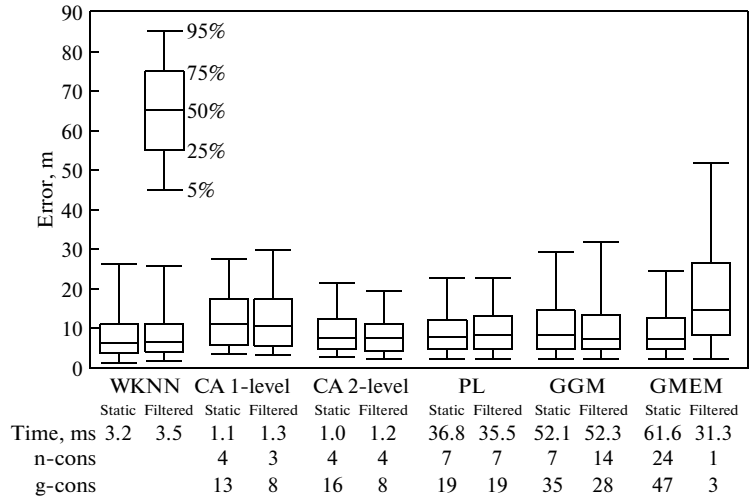


Fig. 6. Method performances with all data.

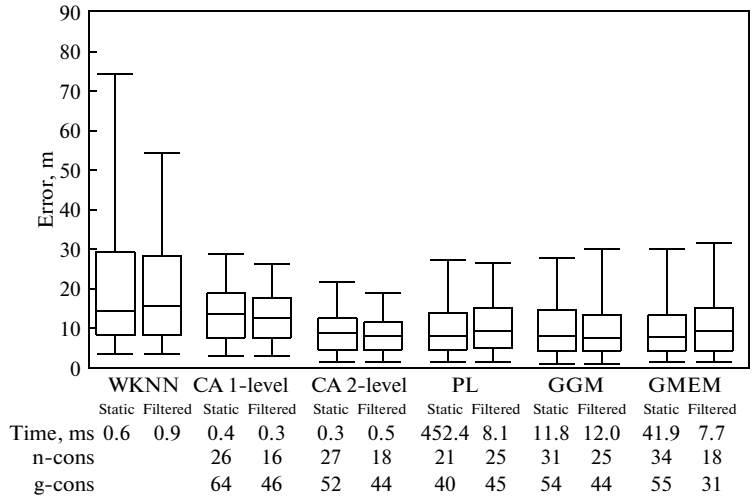


Fig. 7. Method performances with five strongest measurements.

by a GGM with two components. The GMEM modeled the likelihood of the j measurement as

$$p(y_{k,j}|\mathbf{x}_k) = p_{\{\mathrm{N}\}}(f(\mathbf{x}_k)|y_{k,j},6^2), \quad (13)$$

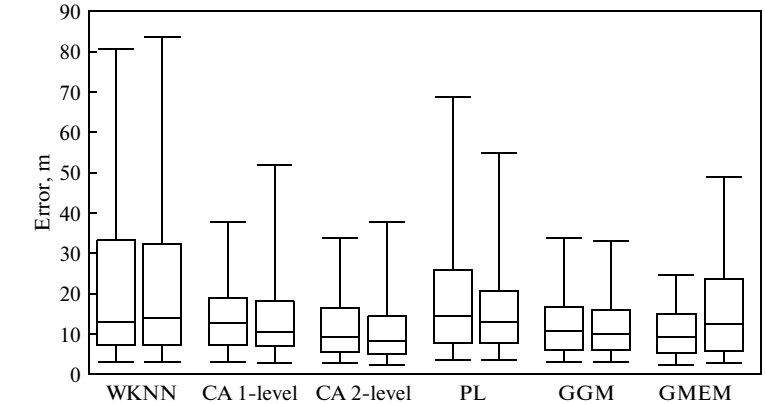
where $p_{\{\mathrm{N}\}}(f(\mathbf{x}_k)|y_{k,j},6^2)$ is the pdf of $\{\mathrm{N}\}(y_{k,j},6^2)$ evaluated at $f(\mathbf{x}_k)$. Function $f(\mathbf{x}_k)$ yields the RSS and is defined as

$$f(\mathbf{x}_k) = \sum_n \omega_{n,\mathrm{ID}_{k,j}} p_{\{\mathrm{N}\}}(\mathbf{x}_k|\boldsymbol{\mu}_{n,\mathrm{ID}_{k,j}},\boldsymbol{\Sigma}_{n,\mathrm{ID}_{k,j}}) - 90\,\mathrm{dBm}, \quad (14)$$

where $\boldsymbol{\mu}_{n,\mathrm{ID}_{k,j}}$ and $\boldsymbol{\Sigma}_{n,\mathrm{ID}_{k,j}}$, are the mean and the covariance matrix of the nth Gaussian component of the GM for the AP with identifier $\mathrm{ID}_{k,j}$ and $\omega_{n,\mathrm{ID}_{k,j}}$ is the component weight.

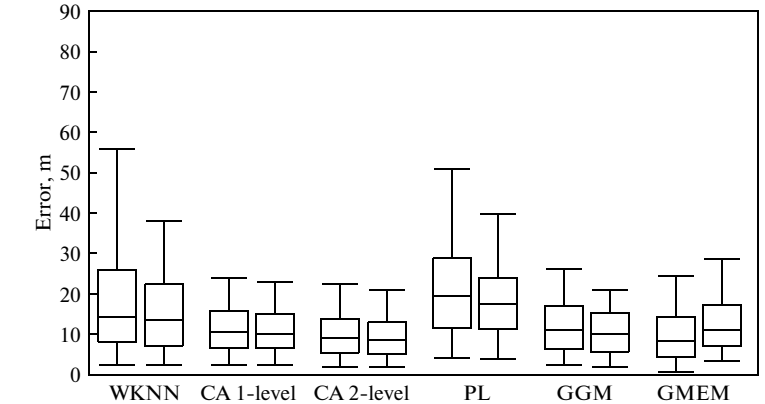
The methods were tested in four different scenarios:

- Fig. 6: full data
- Fig. 7: only the APs with five strongest signals were used for positioning



	WKNN		CA 1-level		CA 2-level		PL		GGM		GMEM	
	Static	Filtered	Static	Filtered	Static	Filtered	Static	Filtered	Static	Filtered	Static	Filtered
Time, ms	0.4	0.6	0.3.	0.4	0.2	0.4	35.4	4.3	3.5	3.7	4.6	3.2
n-cons			37	24	40	31	36	48	50	30	27	35
g-cons			67	53	69	59	60	73	73	61	38	60

Fig. 8. Method performances with 90% of APs dropped.



	WKNN		CA 1-level		CA 2-level		PL		GGM		GMEM	
	Static	Filtered	Static	Filtered	Static	Filtered	Static	Filtered	Static	Filtered	Static	Filtered
Time, ms	0.7	0.9	0.3	0.5	0.3	0.5	417.1	8.4	10.2	10.4	5.5	6.8
n-cons			31	9	42	11	15	10	57	17	13	12
g-cons			62	25	63	26	32	18	88	44	19	23

Fig. 9. Method performances when positioning done several months after data for radio map generation was collected.

• Fig. 8: 90% of APs were dropped pseudorandomly to check how the methods perform in situations with low AP density

• Fig. 9: data for generating the radio maps and data for positioning where collected with a time gap of several months to evaluate the methods' performance degradation over time

In Figs. 6–9 we present quantiles with box plots for the positioning errors, absolute time for one position estimate, and consistency values that can be used to evaluate the accuracy of the estimated position's covariance matrix that is reported by a method. The boxes show the 5%, 25%, 50% (median), 75%, and 95% quantiles of the 2D position errors. For the n-cons (normal consistency [36, p. 235 ff.]) values we assumed Gaussian distributed positioning errors, and computed how often the errors were within the 50% ellipse, i.e.

$$(\hat{\mathbf{x}}_u - \mathbf{x}_u)^T \mathbf{P}_u^{-1} (\hat{\mathbf{x}}_u - \mathbf{x}_u) \leq \chi_2^2(0.5) = 1.3863, \qquad (15)$$

Table 2. Summary of parametric fingerprint methods analyzed in this paper

Method	Radio map entries per CN	Radio map generation method	Positioning method
Coverage area (Gaussian) (1-level) [20, 29]	1 bivariate Gaussian (2 parameters for mean, 3 for covariance)	Bayes' rule using locations where CN's signal is received	Bayes' rule using CA-centres of observed CNs
Coverage area (Gaussian) (2-level) [39]	2 bivariate Gaussians (4 parameters for means, 6 for covariances)	Bayes' rule using locations where CN's signal is received	Bayes' rule using CA-centres of observed CNs
Coverage area (Student-t) (1-level) [37]	1 bivariate Student-t (2 parameters for mean, 3 for shape, 1 for dof)	EM (for MAP) or Gibbs Sampler (posterior) using locations where CN's signal is received	Bayes' rule using CA-centres of observed CNs
Coverage area (Student-t) (2-level) [39]	2 bivariate Student-t:s (4 parameters for mean, 6 for shape, 2 for dof)	EM (for MAP) or Gibbs Sampler (posterior) using locations where CN's signal is received	Bayes' rule using CA-centres of observed CNs
Path loss model Nurminen et al. [50]	2 bivariate Gaussians (1 for CN position, 1 for PLM parameters)	Iterative Reweighted Least Squares (IRLS)	Grid method using standard Monte Carlo, Metropolis-Hastings or IRLS
Path loss model Han et al. [18]	8 parameters (2 for CN position, 6 for PL model)	Solving linear equation using least-squares method	Recreate full FP radio map and WKNN
Path loss model Shrestha et al. [19]	3 parameters for CN position (3-dim.) plus (i) 2 for PL model (8) (ii) 3 for PL model (8) with floor parameter (iii) M for Mth order multi-slope PL model (iv) $M+1$ for multi-slope with floor parameter	MSE minimisation with (a) least-square or (b) weighted least-square or (c) Minimum Mean Square Error	Recreate full FP radio map and KNN ($k=4$)
Generalised Gaussian Mixture [80, 81]	5 parameters (2 for CN, 2 for PL model plus 1 additional)	Iterative Reweighted Least Squares (IRLS)	Bayes' rule using CA-centres of observed CNs and RSS values
Gaussian Mixture by Kaji & Kawaguchi [44]	N bivariate Gaussians for GM components	FP data transformation, parameters GM components fitted by EM	Static problem: grid approach filtering problem: particle filter
GMB-REM [66–68]	1 regularised GM for each grid point	EM for finding maximum log-likelihood of FP data	Multiply likelihood of each grid point with prior pdf to get posterior pdf

where $\hat{\mathbf{x}}_u$ is the estimated UE position, $\mathbf{P}_u^{-1}$ its covariance matrix and $\mathbf{x}_u$ the true UE position. This measure may be used for checking the error estimate in both ways (if it is too small or large) as long as the distribution is close to the normal distribution. In g-cons (general consistency [84]) we computed how often the errors were within 50% for any distribution using the modified Chebyshev inequality, namely

$$(\hat{\mathbf{x}}_u - \mathbf{x}_u)^T \mathbf{P}_u^{-1} (\hat{\mathbf{x}}_u - \mathbf{x}_u) \le \frac{2}{0.5} = 4. \qquad (16)$$

When using all of the data all parametric methods were inconsistent, with n-cons values far from the desirable 50% and g-cons values far from the 50% that can be interpreted as minimum requirement (a g-cons of 60% is not necessarily better/worse than 55%), and there are no significant differences between the accuracies of the different methods, except for filtered GMEM. The computation time for static GMEM is higher than for filtered GMEM because it is computed on a grid, whereas the filtered GMEM uses a particle filter. The results suggest that the 300 particles proposed in [44] was too few. Using only the five strongest measurements improved the consistency and reduced the relative computing time[4] for all methods. Since the GGM's computational demand depends exponentially on the number of measurements [81] the reduction in computation time for static and filtered GGM could be expected, although in our tests the dependence is not exponential due to collapsing a GGM

[4] The large time value for static PL can be explained by the facts that in two (of 308) points the convergence was extremely slow and that our implementation did not restrict the number of iterations.

after five measurements to a single Gaussian. At the same time the positing accuracy degraded significantly only for WKNN and the CA 1-level approach. This is evidence for dependency of the measurements. In the test building there were some MIMO (Multiple Input Multiple Output) APs that produced dependent measurements.

Figure 8 reveals that the more sophisticated approaches (PL, GGM and WKNN) perform worse or similar than the relatively simple CA methods for scenarios with low AP density. The same holds for the scenario in which the radio map was outdated (compare Fig. 9).

One possible reason for the static and the filtered GGM's poor performance in all four scenarios (compared with their performance in [80] and in [81]) might be that we used a different approach for determining the covariance matrices of the GGM's two Gaussian components, since our tests where carried out in a WLAN rather than in a cellular network. A deeper analysis of the GGM can be found in [81].

VIII. CONCLUSION

In this paper we considered different parametric fingerprinting and positioning methods, analyzing weaknesses and strengths. We tested several of those methods with real WLAN data for different test tracks and scenarios; we furthermore compared their positioning accuracy and consistency with each other and a WKNN (as an example of a nonparametric FP method).

Table 2 summarizes the considered parametric methods; what parameters are stored in the radio map instead of FP data, how these parameters are obtained and how they are used for positioning.

All the proposed methods help to significantly reduce the size of the radio map used in the positioning phase, compared with nonparametric methods. In addition, it is possible to update the radio maps used by the CA, PL model, GGM and GMEM approaches we tested as new FPs become available.

Our tests show that all parametric methods, except the CA 1-level and the filtered GMEM method, provide similar positioning accuracy than the nonparametric WKNN in case of a high CN density and when using all available measurements. However, this comes at the cost of significantly higher computation time for the PL model, GGM and GMEM methods. When using only the five strongest measurements their computation time drops significantly. Furthermore, all parametric methods still show similar positioning performances, while the WKNN's performance degrades considerably. This means that the parametric methods need fewer observable CNs than the nonparametric method to achieve satisfying positioning accuracy. When the CN density is low or the mapping data is outdated then the simpler CA 2-level technique achieves at least similar positioning accuracy than the more sophisticated parametric techniques and the WKNN. Thus, the CA technique gives the best trade-off between accuracy and computational demand. The other parametric methods are, like the WKNN, more vulnerable to harsh environments. However, we believe, and studies presented in this paper have shown, that both PL model and GGM approach can outperform the CA methods when thoroughly trained for their specific application, which we excluded from our tests.

It is important to notice that the achieved positioning accuracy of all methods is sufficient for many real-world applications (e.g. weather forecast, advertising), but insufficient for navigation. To improve the methods' performances map information and additional measurements (e.g. from an IMU, AOA or time delay measurements) could be used. We believe that using floor maps would improve their positioning accuracy since, for example, crossing walls could be prohibited. For example, Nurminen et al. [85] show that using the floor map improves their particle filter's positioning accuracy significantly. However, how to combine them with some of the other methods presented in this paper is still an open question. Furthermore, map information can also be used in the offline phase to generate a more accurate radio map.

ACKNOWLEDGMENTS

Philipp Müller acknowledges the financial support of the TUT Doctoral Programme. Matti Raitoharju has been supported by Tampere Doctoral Programme in Information Science and Engineering, Nokia Inc., Nokia Foundation and Jenny and Antti Wihuri Foundation. This research was partly funded by HERE, a Nokia Business.

The authors thank Henri Nurminen for his advice and for provision of Matlab code for his approach [50].

The authors declare that there is no conflict of interests regarding the publication of this article.

REFERENCES

1. Gezici, S., Tian, Z., Biannakis, G.B., Kobayashi, H., Molisch, A.F., Poor, H.V., and Sahinoglu, Z., Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks, *IEEE Signal Processing Magazine*, July 2005, vol. 22, no. 4, pp. 70–84.
2. Gustafsson, F. and Gunnarsson, F., Mobile positioning using wireless networks: possibilities and fundamental limitations based on available wireless network measurements, *IEEE Signal Processing Magazine*, July 2005, vol. 22, no. 4, pp. 41–53.
3. Trevisani, E. and Vitaletti, A., Cell-id location technique, limits and benefits: an experimental study, in *Proceedings of Sixth IEEE Workshop on Mobile Computing Systems and Applications (WMCSA 2004)*, Windermere, Cumbria, United Kingdom, December 2004.
4. Wirola, L., Studies on location technology standards evolution in wireless networks, Ph.D. dissertation, Tampere University of Technology, March 2010,

[Online], Available: http://URN.fi/URN:NBN:fi:tty-201002121065
5. Dardari, D., Conti, A., Ferner, U., Giorgetti, A., and Win, M.Z., Ranging with ultrawide bandwidth signals in multipath environments, *Proceedings of the IEEE*, February 2009, vol. 97, no. 2, pp. 404–426.
6. Mautz, R., Indoor positioning technologies, Habilitation Thesis, http://e-collection.library.ethz.ch/eserv/eth:5659/eth-5659-01.pdf, February 2012.
7. Liu, H., Darabi, H., Banerjee, P., and Liu, J., Survey of wireless indoor positioning techniques and systems, *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, November 2007, vol. 37, no. 6, pp. 1067–1080.
8. Silventoinen, M.I. and Rantalainen, T., Mobile station emergency locating in GSM, in *Proceedings of IEEE International Conference on Personal Wireless Communications*, New Delhi, India, February 1996, pp. 232–238.
9. Constandache, I., Gaonkar, S., Sayler, M., Choudhury, R.R., and Cox, L., EnLoc: Energy-efficient localization for mobile phones, in *2009 Proceedings IEEE INFOCOM*, Rio de Janeiro, Brazil, April 2009, pp. 2716–2720.
10. Spirito, M., Pöykkö, S., and Knuuttila, O., Experimental performance of methods to estimate the location of legacy handsets in GSM, in *IEEE VTS 54th Vehicular Technology Conference (VTC), 2001*, Atlantic City, NJ, USA, October 2001, pp. 2716–2720.
11. el Peral-Rosado, J.A., López-Salcedo, J.A., Seco-Granados, G., Zanier, F., and Crisci, M., Analysis of positioning capabilities of 3GPP LTE, in *Proceedings of 25th International Technical Meeting of the Satellite Division of ION*, Nashville, TN, USA, September 2012.
12. Wirola, L., Laine, T.A., and Syrjärinne, J., Mass-market requirements for indoor positioning and indoor navigation, in *2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Zurich, Switzerland, September 2010.
13. Park, J., Curtis, D., Teller, S., and Ledlie, J., Implications of device diversity for organic localization, in *2011 Proceedings IEEE INFOCOM*, April 2011, pp. 3182–3190.
14. Molisch, A.F., *Wireless Communications*, Wiley - IEEE, January 2011, second edition.
15. Honkavirta, V., Perälä, T., Ali-Löytty, S., and Piché, R., A comparative survey of WLAN location fingerprinting methods, in *Proceedings of the 6th Workshop on Positioning, Navigation and Communication 2009 (WPNC'09)*, Hannover, Germany, March 2009, pp. 243–251.
16. Figuera, C., Mora-Jiménez, I., Guerrero-Curieses, A., Rojo-Álvarez, J.L., Everss, E., Wilby, M., and Ramos-López, J., Nonparametric model comparison and uncertainty evaluation for signal strength indoor location, *IEEE Transactions on Mobile Computing*, September 2009, vol. 8, no. 9, pp. 1250–1264.
17. Müller, P., Raitoharju, M., and Piché, R., A field test of parametric WLAN-fingerprint-positioning methods, in *Proceedings of the 17th International Conference on Information Fusion (Fusion 2014)*, Salamanca, Spain, July 2014, pp. 1–8.
18. Han, S., Gong, Z., Meng, W., and Li, C., An indoor radio propagation model considering angles for WLAN infrastructures, *Wireless Communications and Mobile Computing*, 2015 (in press).
19. Shrestha, S., Talvitie, J., and Lohan, E.S., Deconvolution-based indoor localization with WLAN signals and unknown access point locations, in *2013 International Conference on Localization and GNSS (ICL-GNSS)*, Torino, Italy, June 2013, pp. 1–6.
20. Koski, L., Piché, R., Kaseva, V., Ali-Löytty, S., and Hännikäinen, M., Positioning with coverage area estimates generated from location fingerprints, in *Proceedings of the 7th Workshop on Positioning, Navigation and Communication 2010 (WPNC'10)*, Dresden, Germany, March 2010, pp. 99–106.
21. Machaj, J., Piché, R., and Brida, P., Rank based fingerprinting algorithm for indoor positioning, in *2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Guimaraes, Portugal, September 2011, pp. 6–11.
22. Wirola, L., Wirola, L., and Piché, R., Bandwidth and storage reduction of radio maps for offline WLAN positioning, in *2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Montbeliard-Belfort, France, October 2013.
23. Eisa, S., Peixoto, J., Meneses, F., and Moreira, A., Removing useless APs and fingerprints from WiFi indoor positioning radio maps, in *2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Montbeliard-Belfort, France, 2013, pp. 739–745.
24. Raitoharju, M., Fadjukoff, T., Ali-Löytty, S., and Piché, R., Using unlocated fingerprints in generation of WLAN maps for indoor positioning, in *Proceedings of PLANS 2012 IEEE/ION Position Location and Navigation Symposium*, Myrtle Beach, SC, USA, April 2012, pp. 576–583.
25. Wirola, L., Halivaara, I., and Syrjärinne, J., Requirements for the next generation standardized location technology protocol for location-based services, *Journal of Global Positioning Systems*, 2008, vol. 7, no. 2, pp. 91–103.
26. Kaemarungsi, K and Krishnamurthy, P., Properties of indoor received signal strength for WLAN location fingerprinting, in *The First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (MOBIQUITIOUS 2004)*, Boston, MA, USA, August 2004, pp. 14–23.
27. Vaupel, T., Seitz, J., Kiefer, F., Haimerl, S., and Thielecke, J., Wi-Fi positioning: System considerations and device calibration, in *2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Zurich, Switzerland, September 2010.
28. Laoudias, C., Piché, R., and Panayiotou, C.G., Device signal strength self-calibration using histograms, in *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Sydney, NSW, Australia, November 2012.
29. Koski, L., Perälä, T., and Piché, R., Indoor positioning using WLAN coverage area estimates, in *2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Zurich, Switzerland, September 2010.
30. Haeberlen, A., Flannery, E., Ladd, A.M., Rudys, A., Wallach, D.S., and Kavraki, L.E., Practical robust localization over large-scale 802.11 wireless networks, in *MobileCom'04*, Philadelphia, PA, USA, September-October 2004.
31. Kjaergaard, M.B., Indoor location fingerprinting with heterogeneous clients, *Pervasive and Mobile Computing*, 2011, vol. 7, no. 1, pp. 31–43.

32. Cheng, Y.-C., Chawathe, C., LaMarca, A., and Krumm, A., Accuracy characterization for metropolitan-scale Wi-Fi localization, in *Proceedings of the 3rd International Conference on Mobile Systems, Applications and Services (MobiSys 2005)*, Seattle, WA, USA, June 2005.
33. Hossain, A.M. and Soh, W.-S., Cramér-Rao bound analysis of localization using signal strength difference as location fingerprint, in *2010 Proceedings IEEE INFOCOM*, San Diego, CA, USA, March 2010.
34. Jazwinski, A.H., *Stochastic Processes and Filtering Theory*, ser. Mathematics in Science and Engineering, Academic Press, 1970, vol. 64.
35. Särkkä, S., *Bayesian Filtering and Smoothing*, Cambridge University Press, 2013.
36. Bar-Shalom, Y., Li, R.X., and Kirubarajan, T., *Estimation with Applications to Tracking and Navigation, Theory Algorithms and Software*, John Wiley & Sons, 2001.
37. Piché, R., Robust estimation of a reception region from location fingerprints, in *2011 International Conference on Localization and GNSS (ICL-GNSS)*, Tampere, Finland, June 2011, pp. 31–35.
38. Koski, L., Positioning with Bayesian coverage area estimates and location fingerprints, Master's thesis, Tampere University of Technology, http://math.tut.fi/posgroup/koski_mscth.pdf, March 2010.
39. Raitoharju, M., Ali-Löytty, S., Piché, R., and Dashti, M., Positioning with multilevel coverage area models, in *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Sydney, NSW, Australia, November 2012.
40. Geman, S and Geman, D., Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1984, vol. PAMI-6, no. 6, pp. 721–741.
41. Dempster, A.P., Laird, N.M., and Rubin, D.B., Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.
42. Honkavirta, V., Location fingerprinting methods in wireless local area networks, Master's thesis, Tampere University of Technology, http://urn.fi/URN:NBN:fi:tty-200911137109, November 2008.
43. Laoudias, C., Panayiotou, C.G., and Kemppi, P., On the RBF-based positioning using WLAN signal strength fingerprints, in *Proceedings of the 7th Workshop on Positioning, Navigation and Communication 2010 (WPNC'10)*, Dresden, Germany, March 2010, pp. 93–98.
44. Kaji, K. and Kawaguchi, N., Design and implementation of WiFi indoor localization based on Gaussian mixture model and particle filter, in *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Sydney, NSW, Australia, November 2012.
45. Hashemi, H., The indoor radio propagation channel, *Proceedings of the IEEE*, July 1993, vol. 81, no. 7, pp. 943–968.
46. Patwari, N., Alfred O. Hero, I., Perkins, M., Correal, N.S., and O'Dea, R.J., Relative location estimation in wireless sensor networks, *IEEE Transactions on Signal Processing*, August 2003, vol. 51, no. 8, pp. 2137–2148.
47. Li, X., RSS-based location estimation with unknown pathloss model, *IEEE Transactions on Wireless Communications*, December 2006, vol. 5, no. 12, pp. 3626–3633.
48. Roos, T., Myllymäki, P., and Tirri, H., A statistical modeling approach to location estimation, *IEEE Transactions on Mobile Computing*, Jan-Mar 2002, vol. 1, no. 1, pp. 59–69.
49. Bo-Chieh Liu, S.M., Lin, K.H., and Wu, J.-C., Analysis of hyperbolic and circular positioning algorithms using stationary signal-strength-difference measurements in wireless communications, *IEEE Transactions on Vehicular Technology*, March 2006, vol. 55, no. 2, pp. 499–509.
50. Nurminen, H., Talvitie, J., Ali-Löytty, S., Müller, P., Lohan, E.S., Piché, R., and Renfors, M., Statistical path loss parameter estimation and positioning using RSS measurements in indoor wireless networks, in *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Sydney, NSW, Australia, November 2012.
51. Salo, J., Vuokko, L., El-Sallabi, H.M., and Vainikainen, P., An additive model as a physical basis for shadow fading, *IEEE Transactions on Vehicular Technology*, January 2007, vol. 56, no. 1, pp. 13–26.
52. Erceg, V., Greenstein, L.J., Tjandra, S.Y., Parkoff, S.R., Gupta, A., Kulic, B, Julius, A.A., and Bianchi, R., An empirically based path loss model for wireless channels in suburban environments, *IEEE Journal on Selected Areas in Communications*, 1999, vol. 17, no. 7, pp. 1205–1211.
53. Rappaport, T.S., Seidel, S.Y., and Takamizawa, K., Statistical channel impulse response models for factory and open plan building radio communication system design, *IEEE Transactions on Communications*, 1991, vol. 39, no. 5, pp. 794–807.
54. Ghassenizadeh, S.S., Jana, R., Rice, C.W., Turin, W., and Tarokh, V., A statistical path loss model for in-home UWB channels, in *2002 IEEE Conference on Ultra Wideband Systems and Technologies*, Baltimore, MD, USA, May 2002, pp. 59–64.
55. Dil, B. and Havinga, P., RSS-based localization with different antenna orientations, in *2010 Australasian Telecommunication Networks and Applications Conference (ATNAC)*, Auckland, New Zealand, 2010, pp. 13–18.
56. Shrestha, S., Laitinen, E., Talvitie, J., and Lohan, E.S., RSSI channel effects in cellular and WLAN positioning, in *Proceedings of the 9th Workshop on Positioning, Navigation and Communication 2012 (WPNC'12)*, Dresden, Germany, March 2012, pp. 187–192.
57. Rappaport, T.S., *Wireless Communications - Principles and Practice*, Prentice-Hall, 2010, second edition.
58. Seidel, S.Y. and Rappaport, T.S., 914 MHz path loss prediction models for indoor wireless communications in multifloored buildings, *IEEE Transactions on Antennas and Propagation*, February 1992, vol. 40, no. 2, pp. 207–217.
59. Dil, B and Havinga, P., On the calibration and performance of RSS-based localization methods, in *Internet of Things 2010 Conference*, Tokyo, Japan, November/December 2010.
60. Hata, M., Empirical formula for propagation loss in land mobile radio services, *IEEE Transactions on Vehicular Technology*, August 1980, vol. 29, no. 3, pp. 317–325.
61. Shujaee, K., Ebaid, A., George, R., and Sazegarnejad, M.A., Channel modeling and range

extension for UWB communications using directional antenna in LOS and NLOS path loss models, in *World Automation Congress (WAC 2008)*, Hawaii, HI, USA, September-October 2008.

62. Sorenson, H.W. and Alspach, D.L., Recursive Bayesian estimation using Gaussian sums, *Automatica*, July 1971, vol. 7, no. 4, pp. 465–479.
63. Ali-Löytty, S., Gaussian mixture filters in hybrid positioning, Ph.D. dissertation, Tampere University of Technology, August 2009, Online, Available: http://URN.fi/URN:NBN:fi:tty-200905191055
64. Lo, J. T.-H., Finite-dimensional sensor orbits and optimal nonlinear filtering, *IEEE Transactions on Information Theory*, September 1972, vol. 18, no. 5, pp. 583–588.
65. Alspach, D.L. and Sorenson, H.W., Nonlinear Bayesian estimation using Gaussian sum approximations, *IEEE Transactions on Automatic Control*, August 1972, vol. 17, no. 4, pp. 439–448.
66. Koshizen, T., The sensor selection task of the Gaussians mixture Bayes' with regularized EM (GMB-REM) technique in robot position estimation, in *Proceedings of the 1999 IEEE International Conference on Robotics & Automation*, Detroit, MI Detroit, MI Detroit, MI, USA, May 1999, pp. 2620–2625.
67. Koshizen, T., The evolved Gaussian mixture Bayes' technique using sensor selection task integrated with sensor fusion scheme in mobile robot position estimation, in *Proceedings of 1999 IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA'99)*, Monterey, CA, USA, November 1999, pp. 202–207.
68. Koshizen, T., Bartlett, P., and Zelinsky, A., Sensor fusion of odometry and sonar sensors by the Gaussian mixture Bayes' technique in mobile robot position estimation, in *IEEE SMC '99 Conference Proceedings*, vol. 4, Tokyo, Japan, 1999, pp. 742–747.
69. Ali-Löytty, S. and Sirola, N., Gaussian mixture filter in hybrid navigation, in *Proceedings of The European Navigation Conference GNSS 2007*, Geneva, Switzerland, May 2007, pp. 831–837.
70. Arulampalam, M.S., Maskell, S., Gordon, N., and Clapp, T., A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, *IEEE Transactions on Signal Processing*, 2002, vol. 50, no. 2, pp. 174–188.
71. Ali-Löytty, S., Efficient Gaussian mixture filter for hybrid positioning, in *2008 IEEE/ION Position Location and Navigation Symposium*, Monterey, California, May 2008, pp. 60–66.
72. Ali-Löytty, S. and Sirola, N., Gaussian mixture filter and hybrid positioning, in *Proceedings of ION GNSS 2007*, Fort Worth, Texas, September 2007, pp. 562–570.
73. Raitoharju, M. and Ali-Löytty, S., An adaptive derivative free method for Bayesian posterior approximation, *IEEE Signal Processing Letters*, February 2012, vol. 19, no. 2, pp. 87–90.
74. Anderson, B.D.O. and Moore, J.B., *Optimal Filtering*, ser. Prentice-Hall information and system sciences, Prentice-Hall, 1979.
75. Ali-Löytty, S., On the convergence of the Gaussian mixture filter, Tampere University of Technology, Institute of Mathematics, http://urn.fi/URN:NBN:fi:tty-2011041510704, Tech. Rep. Research report 89, 2008.
76. Salmond, D.J., Mixture reduction algorithms for target tracking, in *IEE Colloquium on State Estimation in Aerospace and Tracking Applications*, London, England, 1989, pp. 7/1–7/4.
77. Schieferdecker, D. and Huber, M.F., Gaussian mixture reduction via clustering, in *Proceedings of the 12th International Conference on Information Fusion (FUSION '09)*, Seattle, WA, USA, 2009, pp. 1536–1543.
78. Crouse, D.F., Willett, P., Pattipati, K., and Svensson, L., A look at Gaussian mixture reduction algorithms, in *Proceedings of the 14th International Conference on Information Fusion (FUSION '11)*, Chicago, IL, USA, 2011.
79. Kristan, M., Skočaj, D., and Leonardis, A., Incremental learning with Gaussian mixture models, in *Computer Vision Winter Workshop 2008*, Moravske Toplice, Slovenia, February 2008, pp. 25–32.
80. Müller, P., Ali-Löytty, S., Dashti, M., Nurminen, H., and Piché, R., Gaussian mixture filter allowing negative weights and its application to positioning using signal strength measurements, in *Proceedings of the 9th Workshop on Positioning, Navigation and Communication 2012 (WPNC'12)*, Dresden, Germany, March 2012, pp. 71–76.
81. Müller, P., Wymeersch, H., and Piché, R., UWB positioning with generalized Gaussian mixture filters, *IEEE Transactions on Mobile Computing*, October 2014, vol. 13, no. 10, pp. 2406–2414.
82. Bishop, C.M. and Svensén, M., Robust Bayesian mixture modeling, in *Proceedings of 12th European Symposium on Artificial Neural Networks (ESANN 2004)*, Bruges, Belgium, April 2004, pp. 69–74.
83. McLachlan, G.J. and Peel, D., Robust cluster analysis via mixtures of multivariate t-distributions, in *Advances in Pattern Recognition*, ser. Lecture Notes in Computer Science, Amin, A., Dori, D., Pudil, P., and Freeman, H., Eds, Springer Berlin Heidelberg, 1998, vol. 1451, pp. 658–666.
84. Ali-Löytty, S., Sirola, N., and Piché, R., Consistency of three Kalman filter extensions in hybrid navigation, in *Proceedings of The European Navigation Conference GNSS 2005*, Munich, Germany, Jul. 2005.
85. Nurminen, H., Ristimäki, A., Ali-Löytty, S., and Piché, R., Particle filter and smoother for indoor localization, in *2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN2013)*, Montbeliard-Belfort, France, October 2013, pp. 137–146.

PUBLICATION 2

Philipp Müller, Simo Ali-Löytty, Marzieh Dashti, Henri Nurminen, and Robert Piché: Gaussian mixture filter allowing negative weights and its application to positioning using signal strength measurements. In *Proceedings of the 2012 9th Workshop on Positioning, Navigation and Communication (WPNC)*, pages 71–76, Dresden, Germany, March 2012.
DOI: 10.1109/WPNC.2012.6268741

Gaussian Mixture Filter Allowing Negative Weights and its Application to Positioning Using Signal Strength Measurements

Philipp MÜLLER, Simo ALI-LÖYTTY, Marzieh DASHTI, Henri NURMINEN and Robert PICHÉ
Tampere University of Technology, Finland
Emails: {philipp.muller, simo.ali-loytty, marzieh.dashti, henri.nurminen, robert.piche}@tut.fi

***Abstract*—This paper proposes a novel Gaussian Mixture Filter (GMF) that allows components with negative weights. In case of a ring-shaped likelihood function, the new filter keeps the number of components low by approximating the likelihood as a Gaussian mixture (GM) of two components, one with positive and the other with negative weight. In this article, the filter is applied to positioning with received signal strength (RSS) based range measurements. The filter is tested using simulated measurements, and the tests indicate that the new GMF outperforms the Extended Kalman Filter (EKF) in both accuracy and consistency.**

I. Introduction

Hybrid navigation means navigation using measurements from different sources, such as Global Navigation Satellite Systems (e.g. GPS), Inertial Measurement Unit, or local wireless networks such as cellular networks, Bluetooth or WLAN. Range, pseudorange, deltarange, altitude, restrictive and compass measurements are examples of measurements in hybrid navigation. This paper focuses on hybrid navigation using different local wireless networks. The ranges from the network's base stations (BS) that are calculated using received signal strengths (RSS) are used as measurements.

Bayesian filtering theory is applied to improve the position and error estimates. The measurement model of range measurements is typically nonlinear. Thus, the traditional Kalman filter cannot be applied. The most popular example of navigation filters is the Extended Kalman Filter (EKF), which linearizes system and measurement models and then applies the Kalman Filter [1, 2]. This filter is commonly used in satellite positioning and has also been applied to hybrid navigation [3].

Unfortunately, the EKF has a serious consistency problem in highly nonlinear situations, which means that it does not work correctly [4]. In highly nonlinear situations sometimes multiple static position solutions exist, which means that the likelihood function has multiple peaks with significant weight. In this case, it is more reasonable to approximate the likelihood as a Gaussian Mixture (GM) and use Gaussian Mixture Filter (GMF) than to approximate it with only one Gaussian as done by the EKF.

One major challenge in using GMF efficiently is keeping the number of components as small as possible without losing significant information. Therefore, this paper introduces a GMF in which the likelihood is approximated as GM consisting of only two Gaussian components. The novel idea that allows to use this small amount of components and simultaneously approximating the multiple peaks of the likelihood function is to bypass the restriction to nonnegative component weights, which for traditional GM has to be fulfilled.

This paper is organized as follows. After problem formulation in section II, the mathematical fundamentals of the new algorithm are presented in section III. Section IV explains how to apply the new filter for navigating and how to select the model parameters. The results and analyses of tests with simulated cellular data are presented in Section V. Section VI concludes.

II. Bayesian Filtering

This section considers the discrete-time non-linear non-Gaussian system

$$x_k = f_{k-1}(x_{k-1}) + w_{k-1}, \tag{1}$$

$$y_k = h_k(x_k) + v_k, \tag{2}$$

where the vectors $x_k \in \mathbb{R}^{n_x}$ and $y_k \in \mathbb{R}^{n_{y_k}}$ represent the state of the system and the measurement at time t_k, $k \in \mathbb{N}$, respectively. The errors w_k and v_k are assumed to be white, mutually independent and independent of the initial state x_0. In the following the density functions of w_k and v_k are denoted by p_{w_k} and p_{v_k}, respectively. The aim of filtering is to find the conditional probability density function (posterior)

$$p(x_k|y_{1:k}),$$

where $y_{1:k} \triangleq \{y_1, \ldots, y_k\}$. The posterior can be determined recursively according to the following relations [5, 6].

Prediction (prior):

$$p(x_k|y_{1:k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|y_{1:k-1})\,\mathrm{d}\,x_{k-1}; \tag{3}$$

Update (posterior):

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k)p(x_k|y_{1:k-1})}{\int p(y_k|x_k)p(x_k|y_{1:k-1})\,\mathrm{d}\,x_k}, \tag{4}$$

where the transition pdf is

$$p(x_k|x_{k-1}) = p_{w_{k-1}}(x_k - f_{k-1}(x_{k-1}))$$

and the likelihood

$$p(y_k|x_k) = p_{v_k}(y_k - h_k(x_k)). \quad (5)$$

The initial condition for the recursion is given by the pdf of the initial state $p(x_0|y_{1:0}) = p(x_0)$. Knowledge of the posterior distribution (4) enables one to compute an optimal state estimate with respect to any criterion. For example, the minimum mean-square error (MMSE) estimate is the conditional mean of x_k [1, 6]. In general and in the case analyzed within this paper, the conditional probability density function cannot be determined analytically. Because of this, there are many approximative solutions of conditional mean. One popular solution is the Extended Kalman Filter (EKF), which applies Kalman filtering to a linerization of system (1), (2).

III. Gaussian Mixture Filter allowing negative weights

In this section an analytic solution to a particular instance of the Bayesian filtering problem (Section II) is presented. Here the motion model (1) is assumed to be linear and Gaussian:

$$x_k = \mathrm{F}_{k-1}x_{k-1} + w_{k-1}, \quad (6)$$

where w_{k-1} is zero-mean Gaussian (Normal) with covariance Q_{k-1}, that is, $w_{k-1} \sim \mathrm{N}(0, \mathrm{Q}_{k-1})$. Moreover the likelihoods are assumed to have forms

$$p(y_k|x_k) = \mathrm{N}_{\mathrm{R}_{k,1}}^{\mathrm{H}_{k,1}x_k}(m_1(y_k)) \left(1 - \bar{c}\,\mathrm{N}_{\mathrm{R}_{k,2}}^{\mathrm{H}_{k,2}x_k}(m_2(y_k))\right) \quad (7)$$

where y_k represents the measurement, $\mathrm{H}_{k,1} \in \mathbb{R}^{n_1 \times n}$, $\mathrm{H}_{k,2} \in \mathbb{R}^{n_2 \times n}$, parameter $\bar{c} = c \cdot (2\pi)^{\frac{n_2}{2}} \sqrt{\det(\mathrm{R}_{k,2})}$, $c \leq 1$, and m_1 and m_2 are some known functions. In the case considered in this paper measurement y is RSS-value from base station BS_i and $m_1(y) = m_2(y)$ are BS_i's position. Function $\mathrm{N}_{\Sigma_j}^{\mu_j}(x)$ denotes the Gaussian density function with vector mean μ_j and covariance matrix Σ_j

$$\mathrm{N}_{\Sigma_j}^{\mu_j}(x) = \frac{\exp\left(-\frac{1}{2}\|x - \mu_j\|^2_{\Sigma_j^{-1}}\right)}{(2\pi)^{\frac{n}{2}}\sqrt{\det(\Sigma_j)}},$$

where $\|x - \mu_j\|^2_{\Sigma_j^{-1}} = (x - \mu_j)^T \Sigma_j^{-1} (x - \mu_j)$. Because of parameter $\bar{c}$ the likelihood (7) is always nonnegative and so it is a valid likelihood. That kind of likelihoods are not entirely new, they have been used in tracking with "negative" information [7]. If $\bar{c} = 0$ and $m_1(y_k) = y_k$ then the conventional Kalman Filter computes the analytic solution. Using likelihood (7) results in a Gaussian mixture, which can also have negative weights. That kind of Gaussian mixture is defined in Definition 1.

Definition 1 (Gaussian Mixture allowing negative weights): An n-dimensional random variable x is an N-component Gaussian Mixture allowing negative weights (GMA) if its probability density function has the form

$$p(x) = \sum_{j=1}^{N} \alpha_j \,\mathrm{N}_{\Sigma_j}^{\mu_j}(x), \quad (8)$$

where $\alpha_j \in \mathbb{R}$ and $\sum_{j=1}^{N} \alpha_j = 1$. Here some weights are possibly negative but it is required that $p(x) \geq 0$ for all $x \in \mathbb{R}^n$. The abbreviation $x \sim \mathrm{M}(\alpha_j, \mu_j, \Sigma_j)_{(j,N)}$ is used.

Mean and covariance formulas of Gaussian Mixture allowing negative weights are the same as for conventional Gaussian Mixture.

Theorem 2 (Mean and covariance of GMA): Let $x \sim \mathrm{M}(\alpha_j, \mu_j, \Sigma_j)_{(j,N)}$ then

$$\mathrm{E}(x) = \sum_{j=1}^{N} \alpha_j \mu_j \stackrel{\triangle}{=} \mu \text{ and}$$

$$\mathrm{V}(x) = \sum_{j=1}^{N} \alpha_j \left(\Sigma_j + (\mu_j - \mu)(\mu_j - \mu)^T\right)$$

Proof:

$$\begin{aligned}\mathrm{E}(x) &= \int x \left(\sum_{j=1}^{N} \alpha_j \,\mathrm{N}_{\Sigma_j}^{\mu_j}(x)\right) \mathrm{d}\,x \\ &= \sum_{j=1}^{N} \alpha_j \int x \,\mathrm{N}_{\Sigma_j}^{\mu_j}(x)\,\mathrm{d}\,x = \sum_{j=1}^{N} \alpha_j \mu_j\end{aligned}$$

$$\begin{aligned}\mathrm{V}(x) &= \int xx^T \left(\sum_{j=1}^{N} \alpha_j \,\mathrm{N}_{\Sigma_j}^{\mu_j}(x)\right) \mathrm{d}\,x - \mu\mu^T \\ &= \sum_{j=1}^{N} \alpha_j \int xx^T \,\mathrm{N}_{\Sigma_j}^{\mu_j}(x)\,\mathrm{d}\,x - \mu\mu^T \\ &= \sum_{j=1}^{N} \alpha_j \left(\Sigma_j + \mu_j\mu_j^T\right) - \mu\mu^T \\ &= \sum_{j=1}^{N} \alpha_j \left(\Sigma_j + (\mu_j - \mu)(\mu_j - \mu)^T\right)\end{aligned}$$

■

In contrast to the new Gaussian Mixture Filter (GMFA) (Algorithm 1), conventional Gaussian Mixture Filter (GMF) does not allow negative weights [8, 9].

Algorithm 1 Gaussian mixture filter allowing negative weights

Initial state at time t_0: $x_0 \sim \mathrm{M}(\alpha_{i,0}, \mu_{i,0}, \Sigma_{i,0})_{(i,n_0)}$

for $k = 1$ to n_{meas} **do**

1) Prediction (see Sec. III-A):

$$x_k|y_{1:k-1} \sim \mathrm{M}(\alpha_{i,k}^-, \mu_{i,k}^-, \Sigma_{i,k}^-)_{(i,n_k^-)}$$

2) Update (see Sec. III-B):

$$x_k|y_{1:k} \sim \mathrm{M}(\alpha_{i*j,k}, \mu_{i*j,k}, \Sigma_{i*j,k})_{(i*j,\bar{n}_k)}$$

3) Re-indexing and possibly reducing the number of components (see Sec. III-C):

$$x_k|y_{1:k} \sim \mathrm{M}(\alpha_{i,k}, \mu_{i,k}, \Sigma_{i,k})_{(i,n_k)}$$

end for

A. Prediction, Step (1)

Prediction is based on Eq. (3). Previous posterior distribution is

$$x_{k-1}|y_{1:k-1} \sim \mathrm{M}(\alpha_{i,k-1}, \mu_{i,k-1}, \Sigma_{i,k-1})_{(i,n_{k-1})}.$$

$$\begin{aligned} p(x_k|y_{1:k-1}) &= \int p(x_k|\xi) p_{x_{k-1}|y_{1:k-1}}(\xi)\,\mathrm{d}\,\xi \\ &= \int p(x_k|\xi) \left(\sum_{i=1}^{n_{k-1}} \alpha_{i,k-1}\, \mathrm{N}^{\mu_{i,k-1}}_{\Sigma_{i,k-1}}(\xi) \right) \mathrm{d}\,\xi \\ &= \sum_{i=1}^{n_{k-1}} \alpha_{i,k-1} \int p(x_k|\xi)\, \mathrm{N}^{\mu_{i,k-1}}_{\Sigma_{i,k-1}}(\xi)\,\mathrm{d}\,\xi \\ &= \sum_{i=1}^{n_{k-1}} \alpha_{i,k-1}\, \mathrm{N}^{\mathrm{F}_{k-1}\mu_{i,k-1}}_{\mathrm{F}_{k-1}\Sigma_{i,k-1}\mathrm{F}^T_{k-1}+\mathrm{Q}_{k-1}}(x_k) \end{aligned} \tag{9}$$

$$x_k|y_{1:k-1} \sim \mathrm{M}(\alpha^-_{i,k}, \mu^-_{i,k}, \Sigma^-_{i,k})_{(i,n^-_k)},$$

where

$$\begin{aligned} n^-_k &= n_{k-1}, \\ \alpha^-_{i,k} &= \alpha_{i,k-1}, \\ \mu^-_{i,k} &= \mathrm{F}_{k-1}\mu_{i,k-1} \text{ and} \\ \Sigma^-_{i,k} &= \mathrm{F}_{k-1}\Sigma_{i,k-1}{\mathrm{F}_{k-1}}^T + \mathrm{Q}_{k-1}. \end{aligned}$$

Since the previous posterior and $p(x_k|\xi)$ are always non-negative the prior distribution $p(x_k|y_{1:k-1})$ is always non-negative and therefore a valid GMA.

B. Update, Step 2

The update is based on Eq. (4) and likelihood is given in (7).

$$\begin{aligned} &p(x|y_{1:k}) \propto p(y_k|x)p(x|y_{1:k-1}) = p(y_k|x) \sum_{i=1}^{n_k} \alpha^-_{i,k}\, \mathrm{N}^{\mu^-_{i,k}}_{\Sigma^-_{i,k}}(x) \\ &= \sum_{i=1}^{n_k} \alpha^-_{i,k} \left(\mathrm{N}^{\mathrm{H}_1 x}_{\mathrm{R}_{k,1}}(m_1) \left(1 - \bar{c}\, \mathrm{N}^{\mathrm{H}_2 x}_{\mathrm{R}_{k,2}}(m_2)\right) \mathrm{N}^{\mu^-_{i,k}}_{\Sigma^-_{i,k}}(x) \right) \\ &= \sum_{i=1}^{n_k} \alpha^-_{i,k}\, \mathrm{N}^{\mathrm{H}_1\mu^-_{i,k}}_{\bar{\mathrm{R}}_{1,i}}(m_1) \left(\mathrm{N}^{\mu_{i*1,k}}_{\Sigma_{i*1,k}}(x) \left(1 - \bar{c}\, \mathrm{N}^{\mathrm{H}_2 x}_{\mathrm{R}_{k,2}}(m_2)\right)\right) \\ &= \sum_{i=1}^{n_k} \alpha^-_{i,k}\, \mathrm{N}^{\mathrm{H}_1\mu^-_{i,k}}_{\bar{\mathrm{R}}_{1,i}}(m_1) \cdot \\ &\qquad \left(\mathrm{N}^{\mu_{i*1,k}}_{\Sigma_{i*1,k}}(x) - \bar{c}\, \mathrm{N}^{\mu_{i*2,k}}_{\Sigma_{i*2,k}}(x)\, \mathrm{N}^{\mathrm{H}_2\bar{\mu}_{1,i}}_{\bar{\mathrm{R}}_{2,i}}(m_2) \right) \end{aligned} \tag{10}$$

where

$$\begin{aligned} \bar{n}_k &= 2n^-_k, \\ \alpha_{i*1,k} &\propto \alpha^-_{i,k}\, \mathrm{N}^{\mathrm{H}_1\mu^-_{i,k}}_{\bar{\mathrm{R}}_{1,i}}(m_1), \\ \alpha_{i*2,k} &\propto -\bar{c}\alpha^-_{i,k}\, \mathrm{N}^{\mathrm{H}_1\mu^-_{i,k}}_{\bar{\mathrm{R}}_{1,i}}(m_1)\, \mathrm{N}^{\mathrm{H}_2\mu_{i*1,k}}_{\bar{\mathrm{R}}_{2,i}}(m_2), \\ \mu_{i*1,k} &= \mu^-_{i,k} + \mathrm{K}_{i*1,k}(m_1 - \mathrm{H}_1\mu^-_{i,k}), \\ \mu_{i*2,k} &= \mu_{i*1,k} + \mathrm{K}_{i*2,k}(m_2 - \mathrm{H}_2\mu_{i*1,k}), \\ \Sigma_{i*1,k} &= (\mathrm{I} - \mathrm{K}_{i*1,k}\mathrm{H}_1)\Sigma^-_{i,k} \\ \Sigma_{i*2,k} &= (\mathrm{I} - \mathrm{K}_{i*2,k}\mathrm{H}_2)\Sigma_{i*1,k} \\ \mathrm{K}_{i*1,k} &= \Sigma^-_{i,k}\mathrm{H}_1^T \left(\mathrm{H}_1\Sigma^-_{i,k}\mathrm{H}_1^T + \mathrm{R}_{k,1}\right)^{-1} \\ \mathrm{K}_{i*2,k} &= \Sigma_{i*1,k}\mathrm{H}_2^T \left(\mathrm{H}_2\Sigma_{i*1,k}\mathrm{H}_2^T + \mathrm{R}_{k,2}\right)^{-1} \\ \bar{\mathrm{R}}_{1,i} &= \mathrm{H}_1\Sigma^-_{i,k}\mathrm{H}_1^T + \mathrm{R}_{k,1} \text{ and} \\ \bar{\mathrm{R}}_{2,i} &= \mathrm{H}_2\Sigma_{i*1,k}\mathrm{H}_2^T + \mathrm{R}_{k,2}. \end{aligned}$$

For simplicity, abbreviations $x = x_k$, $m_1 = m_1(y_k)$, $m_2 = m_2(y_k)$, $\mathrm{H}_1 = \mathrm{H}_{k,1}$ and $\mathrm{H}_2 = \mathrm{H}_{k,2}$ are used. Since prior and likelihood are always non-negative the posterior is always non-negative and therefore a valid GMA.

C. Re-indexing and possible reduce the number of components, Step 3

In this step the posterior is re-indexed to obtain

$$x_k|y_{1:k} \sim \mathrm{M}(\alpha_{i,k}, \mu_{i,k}, \Sigma_{i,k})_{(i,n_k)}$$

If the number of components is not reduced then $n_k = \bar{n}_k$ and the posterior is the analytic solution of the Bayesian filtering problem.

However, if the number of components should be reduced, the conventional methods are forgetting and merging [8–11]. When these methods are applied to GMFA it has to be ensured that the resulting Gaussian mixture is a valid probability density function. One possible approach would be always forgetting negative weights and/or collapsing the whole posterior to one Gaussian. The latter is used in this paper, in Section V.

IV. How to select parameters for application

In this section the GMFA is applied for navigation based on wireless radio signals. It is assumed that there is a network of base stations that transmit signals with their specific transmission power. The user is able to identify the transmitting BS and to measure the RSS level. RSS level is a base-10 logarithm of the signal intensity, so the simple path loss model

$$y = a - 10n\log_{10}(||x_{\mathrm{bs}} - r_u||) + w, \tag{11}$$

is used to model the measurement. There y represents the RSS level, a, n and x_{bs} are model parameters and $w \sim \mathrm{N}(0, \mathrm{R})$ is the stochastic noise term. a is the RSS level in the unit range, n is the path loss exponent, and x_{bs} is the BS-position. Variable r_u represents the position of the user equipment (UE).

The model parameters are assumed to be known, and the coverage area of the BS is modeled as a Gaussian ellipse using

the algorithm proposed in [12, 13]. This approach has the advantage that it requires a significantly smaller database than techniques that rely on fingerprint data for both, radio map generation and navigation.

The model (11) produces a ring-shaped likelihood with BS-position x_{bs} as the centre point. In this paper, this distribution is approximated by the difference of two Gaussian components. They are both x_{bs}-centered (i.e. $\mathrm{H}_{k,2}x_k = x_{\text{bs}}$ and $\mathrm{H}_{k,1}x_k = x_{\text{bs}}$) and have covariance matrices $\mathrm{R}_{k,1} = \sigma^2_{\max}\mathrm{I}$ and $\mathrm{R}_{k,2} = \sigma^2_{\min}\mathrm{I}$ with $\sigma_{\min} < \sigma_{\max}$. Component 1 has positive and component 2 negative weight. Parameter c in (7) is set to 1, which ensures that the likelihood is nonnegative.

The exact likelihood function $p(y|r_u)$ gets its maximum value when $w = 0$, the range being

$$r \triangleq ||x_{\text{bs}} - r_u|| = 10^{\frac{y-a}{-10n}}, \tag{12}$$

so inequalities $\sigma_{\min} \leq r$ and $\sigma_{\max} > r$ should hold. The derivative of r with respect to y in equation (12) increases as y decreases, so the variance of the range increases as y decreases and the tails of the distribution are heavier outside the ring than inside. Therefore, the approximative model

$$\sigma_{\min} = \max\{1, 0.68r - 48\} \tag{13}$$

$$\sigma_{\max} = 0.9r + 23, \tag{14}$$

is adopted. Equations (13) and (14) are both of form $\sigma = \alpha r + \beta$, and α, β are optimized so that the Lissack-Fu distance [9], which is defined as $\int |p_{\text{true}}(y|r_u) - p_{\text{approx}}(y|r_u)|\,\mathrm{d}\,r_u$, is minimized in the off-line phase.

Figure 1 shows some examples for the performance of the likelihood function (11). The upper row contains plots of likelihoods for strong (left), medium (middle) and weak RSS-values for which exact value of $p(y_k|r_{u,k})$ is calculated in points of a grid with step width 100 and using $a = 30$, $n = 3.5$, $x_{bs} = [0, 0]^T$ and $\mathrm{R} = 36$. In the lower row there are the likelihoods derived from function (11). The errors of the likelihood, which are normalized in the grid area, are 0.2118 (strong RSS), 0.1117 (medium RSS) and 0.1002 (weak RSS).

V. Simulations

For evaluating the performance of the proposed GMFA, in comparison to EKF and a coverage area algorithm (CAF, [14]), simulations with cellular data were done using Matlab. In contrast with GMFA, the CAF uses only coverage areas for positioning, and ignores RSS values.

The state $x = \begin{bmatrix} r_u \\ v_u \end{bmatrix}$ consists of user position vector r_u and user velocity vector v_u, which are in East-North coordinate system; the Up-coordinate is neglected. Motion model matrix $\mathrm{F}_{k-1} = \begin{bmatrix} \mathrm{I} & \Delta t\mathrm{I} \\ 0 & \frac{9}{10}\mathrm{I} \end{bmatrix}$, which means that constant velocity is used (damped with constant $\frac{9}{10}$), and $w_{k-1} \sim \mathrm{N}(0, \mathrm{Q})$ with $\mathrm{Q} = 9\begin{bmatrix} \frac{(\Delta t)^3}{3}\mathrm{I} & \frac{(\Delta t)^2}{2}\mathrm{I} \\ \frac{(\Delta t)^2}{2}\mathrm{I} & \Delta t\mathrm{I} \end{bmatrix}$. For the simulations $\Delta t = 1$.

One-hundred different test tracks are simulated. In the simulations, RSS measurements are used, calculated according to equation (12). Therefore, for each simulation (i.e. test track) 3,500 base station positions (results in table 1) or 10,000 (results in table 2) respectively are simulated as 2-dimensional vectors, uniformly distributed on a 15-by-15 km square around the simulated track. Path loss model parameter a is modeled as $a \sim \mathrm{N}(0, 18^2)$ for each base station, and path loss exponent n is modeled as $n \sim \mathrm{N}(3, 0.7^2)$ with the constraint $n \geq 2$. This choice for distribution of n is based on the fact that the path loss exponent for normal environments is in the range between 2 and 4, and $n = 2$ when the signal propagates through free space [15]. The shadowing variance is fixed to $\mathrm{R} = 36$.

In the simulation, the size of cell coverage areas is chosen based on earlier studies and our experimental knowledge. The size of a cell may range from some meters to some kilometers, for instance 0.1-1 km is reported for microcells [16, 17]. In [16], the experimental results of cell-ID location technique are presented. The average distance between cell-ID location estimate and GPS location estimate is reported to be 800 meters in USA and 500 meters in Italy. For the purpose of this paper, semi-major axis e_1 and semi-minor axis e_2 of coverage area ellipses are therefore simulated as $e_1, e_2 \sim \mathrm{N}(650, 500^2)$, centers of the ellipses are simulated as $c_e \sim \mathrm{N}(x_{\text{bs}}, 200^2\mathrm{I})$, and ellipse angles γ_e are chosen from a continuous uniform distribution on the interval $[0, \pi)$.

Table 1 contains the summary of hundred 300 second simulations using CAF [12–14], EKF and GMFA that is modified such that it approximates the posterior as one Gaussian at every time step in order to increase calculation speed. The simulations are forced to a very poor geometry, at each time step only one base station measurement at most is available. Therefore, if more than one BS are heard then all of them except for one randomly chosen are discarded. In addition, this measurement is used in the next n_{rm} steps, where n_{rm} is a discrete uniform random number between 1 and 10. This emulates real world cases, in which not always new measurements could be obtained every second. Note that it is also possible that any BS is heard outside its coverage area (limited to 1.5 times the coverage area). For the three approaches results using static positioning and filtering are presented. For filtering of CAF the basic Kalman filter is applied. In the static case for EKF, firstly the coverage areas are processed in order to get a position estimate. This estimate is then used as prior for the EKF.

The columns of table 1 are as follow: *Time* is relative computation time using Matlab and the specific implementation, so that computation time of EKF is one. This gives a rough idea of each algorithm's complexity. *Mean* is mean and *Med.* is median, and 95% *err.* is 95th percentile of the 2D position error. Column *Cons.* $\%$ displays the percentage of time steps for which the filters are consistent with respect to the Gaussian consistency test, with risk level 5%.

For the static case GMFA clearly outperforms EKF in terms of consistency, and reduces the three error types significantly. These improvements come with the cost of an approximately 70% larger calculation time. In comparison to the CAF the GMFA provides only modest improvements of mean and

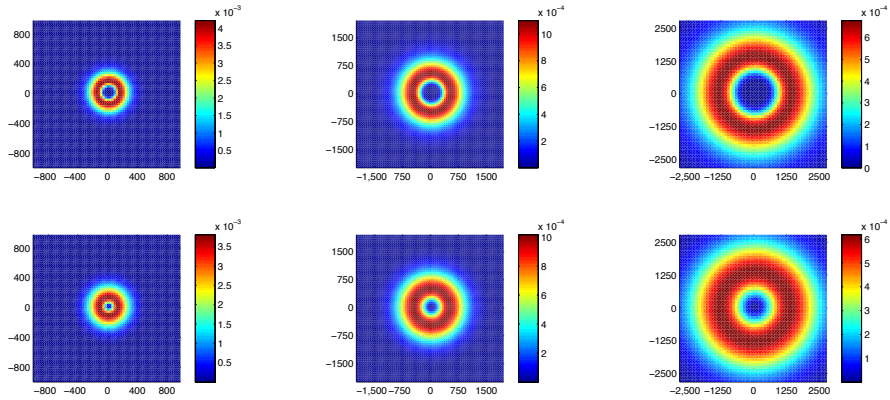


Figure 1. **Approximation of ring-shaped likelihood:** In upper row are plots of likelihoods for strong (left), medium (middle) and weak RSS-values for which exact value of $p(y_k|x_k)$ is calculated in points of a grid with step width 30 (left), 50 (middle) and 100. The lower row contains plots of the likelihoods derived from function (11).

median error, which are unable to justify the major increase in calculation time. Contrary, for the filtered case the improvements of GMFA, compared with CAF, in terms of the three error statistics are justifying the significant increase in calculation time. Compared with the EKF, the GMFA provides similar decreases in the errors as for the static case, but the advantage of significantly higher consistency does not hold.

The weak overall consistencies might be a result of using the Gaussian consistency test that assumes Gaussian distributions, which does not hold in the analyzed cases. An alternative, which should ensure higher consistency levels, is the general inconsistency test, introduced in [4].

Solver	time	Mean	Med	95% err	cons. %
CAF (static)	0.94	702	643	1445	86
EKF (static)	1	758	653	1712	33
GMFA (static)	1.69	670	597	1452	84
CAF (filtered)	0.75	625	572	1304	37
EKF (filtered)	1	477	376	1250	31
GMFA (filtered)	1.75	431	345	1087	38

Table I

Table 1: Summary of 100 different simulations with very poor geometry. Simulations use only one base station measurement at most per time step and usually measurement is excatly the same than previous time step. Time of EKF is used as reference time.

In Table 2 the summary of hundred simulations using CAF, EKF and GMFA are listed. For those tests a good geometry was assumed, at each time step up to six base station measurements are available. It becomes visible that GMFA outperforms CAF and EKF also for good geometries. However, the improvement in the filtered case compared with EKF is less significant and comes at the cost of stronger increase in calculation time than for very poor geometry.

The reason for the significantly larger calculation time of GMFA, for the good geometry case, compared with both other methods is that it uses a maximum of $2^6 = 64$ components, whereas CAF and EKF use only one. In the filtered case, the whole posterior is collapsed to one Gaussian and used as prior in the next time step. Thus, the maximum number of components is still 64, and therefore the run time of GMFA for both static and filtered case are almost similar, with respect to the EKF for the corresponding case. Note that the algorithm is nevertheless clearly faster than common Gaussian mixture filters, which require non-negative component weights. Within these GMFs the likelihood is, mainly for non-linear cases, typically approximated by a large number of Gaussians [9].

Solver	time	Mean	Med	95% err	cons. %
CAF (static)	1.01	423	368	933	61
EKF (static)	1	328	253	857	43
GMFA (static)	14.63	248	184	674	70
CAF (filtered)	0.78	246	224	508	37
EKF (filtered)	1	112	84	290	55
GMFA (filtered)	16.24	100	84	232	65

Table II

Table 2: Summary of 100 different simulations with good geometry. Simulations use up to six base station measurements per time step. Time of EKF is used as reference time.

VI. Conclusion

In this article, a Gaussian mixture filter (GMF) for hybrid positioning applications has been studied. A new way for approximating the likelihood function as a Gaussian mixture with only two components, one having negative weight, was introduced (GMFA). This allows using the GMF (more) efficiently (than conventional GMFs) by keeping the number of components small without losing significant information. In both very poor and good geometry, the new filter (GMFA) clearly outperform the EKF, which has serious consistency problems in highly nonlinear situations that were analyzed

in the paper. It also provides better results than the CAF, since in addition to the coverage areas of base stations, it also uses RSS-measurements, which are generally available in the UE. For real-world applications a tuning strategy should be applied to improve the computational time. In the future, it will be analyzed if using the GMFA only in highly nonlinear cases, and simpler approaches for almost linear cases, results in insignificant accuracy loose and major drop in calculation time. In addition, it will be studied more deeply how the GMFA behaves for cases in which the likelihood displays a shape other than the ring-shape examined in this paper.

ACKNOWLEDGMENTS

This research was funded by Nokia Corporation.

REFERENCES

[1] Y. Bar-Shalom, R. X. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation, Theory Algorithms and Software*. John Wiley & Sons, 2001.

[2] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, ser. Mathematics in Science and Engineering. Academic Press, 1970, vol. 64.

[3] C. Ma, "Integration of GPS and cellular networks to improve wireless location performance," *Proceedings of ION GPS/GNSS 2003*, pp. 1585–1596, 2003.

[4] S. Ali-Löytty, N. Sirola, and R. Piché, "Consistency of three Kalman filter extensions in hybrid navigation," in *Proceedings of The European Navigation Conference GNSS 2005*, Munich, Germany, Jul. 2005.

[5] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, ser. Statistics for Engineering and Information Science. Springer, 2001.

[6] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter, Particle Filters for Tracking Applications*. Boston, London: Artech House, 2004.

[7] W. Koch, "On 'negative' information in tracking and sensor data fusion: Discussion of selected examples," in *Proceedings of the Seventh International Conference on Information Fusion*, P. Svensson and J. Schubert, Eds., vol. I. Mountain View, CA: International Society of Information Fusion, Jun 2004, pp. 91–98. [Online]. Available: http://www.fusion2004.foi.se/papers/IF04-0091.pdf

[8] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, no. 4, pp. 465–479, July 1971.

[9] S. Ali-Löytty, "Gaussian mixture filters in hybrid positioning," Ph.D. dissertation, Tampere University of Technology, August 2009. [Online]. Available: http://URN.fi/URN:NBN:fi:tty-200905191055

[10] S. Ali-Löytty and N. Sirola, "Gaussian mixture filter in hybrid navigation," in *Proceedings of The European Navigation Conference GNSS 2007*, May 2007, pp. 831–837.

[11] D. J. Salmond, "Mixture reduction algorithms for target tracking," *State Estimation in Aerospace and Tracking Applications, IEE Colloquium on*, pp. 7/1–7/4, 1989.

[12] L. Koski, R. Piché, V. Kaseva, S. Ali-Löytty, and M. Hännikäinen, "Positioning with coverage area estimates generated from location fingerprints," in *Proceedings of the 7th Workshop on Positioning, Navigation and Communication 2010 (WPNC'10)*, March 2010, pp. 99–106.

[13] L. Koski, T. Perälä, and R. Piché, "Indoor positioning using wlan coverage area estimates," in *2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, September 2010, pp. 99–106.

[14] M. Dashti, S. Ali-Löytty, L. Wirola, P. Müller, H. Nurminen, and R. Piché, "Robust Kalman filter for positioning with coverage areas of wireless bss," submitted to WPNC 2012.

[15] A. F. Molisch, *Wireless Communications*, 2nd ed. Wiley - IEEE, January 2011.

[16] E. Trevisani and A. Vitaletti, "Cell-id location technique, limits and benefits: an experimental study," in *Proceedings of Sixth IEEE Workshop on Mobile Computing Systems and Applications (WMCSA 2004)*. Windermere, Cumbria, United Kingdom, December 2004.

[17] R. Krievs, "Using fading to improve accuracy of cell id based mobile positioning algorithms: analysis of special cases," in *Scientific Proceedings of RTU. Series 7. Telecommunications and Electronics*, 2002, pp. 50–58.

PUBLICATION 3

Philipp Müller, Henk Wymeersch, and Robert Piché: UWB positioning with generalized Gaussian mixture filters. *IEEE Transactions on Mobile Computing*, Vol. 13, No. 10, pages 2406–2414, October 2014.
DOI: 10.1109/TMC.2014.2307301

UWB Positioning with Generalized Gaussian Mixture Filters

Philipp Müller, Henk Wymeersch, *Member, IEEE,* and Robert Piché, *Member, IEEE*

Abstract—Low-complexity Bayesian filtering for nonlinear models is challenging. Approximative methods based on Gaussian mixtures (GM) and particle filters are able to capture multimodality, but suffer from high computational demand. In this paper, we provide an in-depth analysis of a generalized GM (GGM), which allows component weights to be negative, and requires significantly fewer components than the traditional GM for ranging models. Based on simulations and tests with real data from a network of UWB nodes, we show how the algorithm's accuracy depends on the uncertainty of the measurements. For nonlinear ranging the GGM filter outperforms the extended Kalman filter (EKF) in both positioning accuracy and consistency in environments with uncertain measurements, and requires only slightly higher computational effort when the number of measurement channels is small. In networks with highly reliable measurements, the GGM filter yields similar accuracy and better consistency than the EKF.

Index Terms—Bayesian filtering, Gaussian mixture, Indoor positioning, UWB

1 Introduction

Location-aware applications are enabled by positioning techniques, and are an essential feature of many commercial, public service, and military wireless networks [1], [2]. An important requirement for many applications is that they allow positioning in real time using limited energy resources and computational effort, enabling local processing on small mobile devices.

In outdoor environments, localization and navigation techniques mainly rely on Global Navigation Satellite System (GNSS) signals. However, indoors and also under forest canopies and in certain urban settings, such as urban canyons, poor signal penetration by GNSS generally results in unavailable or unreliable location information [3]. Therefore, positioning in those environments must rely on other measurements, e.g. from an inertial measurement unit (IMU) or from radio signals such as cellular networks, Bluetooth, wireless local area networks (WLAN), or ultra-wideband (UWB). In particular, UWB has attracted a great deal of interest [1], [4], [5], [6], as the use of extremely large bandwidths enables accurate range estimates and high reliability [1].

- *P. Müller and R. Piché are with the Department of Automation Science and Engineering (ASE), Tampere University of Technology, P.O. Box 692, FI-33101 Tampere, Finland. E-mail: philipp.muller, robert.piche@tut.fi*
- *H. Wymeersch is with the Department of Signals and Systems, Chalmers University of Technology, S-41296 Gothenburg, Sweden. E-mail: henkw@chalmers.se*

Philipp Müller acknowledges the financial support of the Doctoral Programme of TUT's President, and the financial support by the Finnish Doctoral Programme in Computational Sciences (FICS) and TUT-Foundation (TTY-säätiö) for his research visit to Chalmers University of Technology, Sweden. This research was supported, in part, by the European Research Council, under Grant No. 258418 (COOPNET).

To track movements in UWB networks, Bayesian filters are commonly employed [7], [8], [9]. For linear Gaussian models, the Kalman filter is optimal [7, p. 206 ff.], while in mildly nonlinear conditions the extended Kalman filter (EKF) is a popular choice [10], [11]. In contrast to these low-complexity methods, in highly nonlinear conditions computationally demanding methods such as the particle filter are required to achieve high accuracy [12]. However, for small mobile devices the particle filter is not an attractive solution due to its computational requirements. The EKF can also be employed in highly nonlinear conditions, but suffers from a number of drawbacks. In particular, the EKF can seriously underestimate the posterior covariance in highly nonlinear situations [13]. In nonlinear situations, the likelihood function can have multiple peaks. Those peaks could be captured by approximating the likelihood as a Gaussian mixture (GM) and using a Gaussian mixture filter (GMF). To reduce the complexity of the GMF, it is important to keep the number of Gaussian components as small as possible without losing significant information. With this tradeoff in mind, a generalized version of Gaussian mixture (GGM) that relaxes the non-negativity restriction on component weights has been proposed [14]. This relaxation makes the GGM more flexible and, under certain circumstances, leads to a significant reduction in the number of Gaussian components. For localization in cellular networks, it has been found in [14] that the GGM filter (GGMF) outperforms the EKF in terms of accuracy and consistency.

In this paper, we build on the work from [14], and perform an in-depth analysis of the GGMF. We study under which circumstances the GGMF

yields satisfying position accuracy, and how its parameters affect its approximation quality for the exact likelihood function. Furthermore, we examine the GGMF's computational complexity. For evaluation purposes we performed extensive simulations and tests with real data in an UWB network. The algorithm's positioning performance and complexity are compared with those of the EKF to see how the algorithms cope with highly nonlinear measurements.

This paper is organized as follows. After the problem formulation in Section 2, the mathematical fundamentals of EKF and GGM/GGMF are presented in Section 3, and their computational complexity is analyzed. Section 4 describes how to apply GGM/GGMF for isotropic Gaussian ranging models. The results and analyses of tests with simulated and measured UWB data are presented in Section 5. Section 6 concludes the paper.

Notation : We will denote by $\mathcal{N}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})$ a Gaussian density function with vector mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The expectation and covariance operators are written as $\mathrm{E}(\cdot)$ and $\mathrm{V}(\cdot)$, respectively. Vectors will be denoted in bold (e.g., $\mathbf{x}$) and matrices in bold capitals (e.g., $\mathbf{X}$).

2 System model

In Sections 4 and 5 we will focus on nonlinear ranging. Therefore, we consider the discrete-time nonlinear system

$$\mathbf{x}_k = \mathbf{F}_k\mathbf{x}_{k-1} + \mathbf{w}_k, \tag{1a}$$

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k, \tag{1b}$$

where the vectors $\mathbf{x}_k \in \mathbb{R}^{n_\mathbf{x}}$ and $\mathbf{y}_k \in \mathbb{R}^{n_{\mathbf{y}_k}}$ represent the state of the system and the vector of independent range measurements at time t_k, $k \in \mathbb{N}$, respectively. Examples of range measurements include received-signal-strength-based ranging [14] and time-of-arrival-based ranging [1]. The function $\mathbf{h}_k(\mathbf{x}_k)$ returns a vector of Euclidean distances between the position contained in state $\mathbf{x}_k$ and the locations $\mathbf{x}_{\mathrm{AP}}$ of reference nodes/access points (APs).

In the following we denote the probability density functions (pdf) of the errors $\mathbf{w}_k$ and $\mathbf{v}_k$ by $p_{\mathbf{w}_k}(\cdot)$ and $p_{\mathbf{v}_k}(\cdot)$, respectively. The aim of Bayesian filtering is, at every time t_k, to determine the posterior pdf $p(\mathbf{x}_k|\mathbf{y}_{1:k})$, where $\mathbf{y}_{1:k} \triangleq [\mathbf{y}_1,\ldots,\mathbf{y}_k]$ denotes the measurement history. Assuming[1] $\mathbf{w}_k$ and $\mathbf{v}_k$ to be white, mutually independent, with covariance matrices $\mathbf{Q}_k$ and diagonal (due to assumption of independent range measurements) $\mathbf{R}_k$ respectively, and independent of the initial state $\mathbf{x}_0$, the posterior can be determined recursively according to the following relations [11]:
Prediction (prior update):

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})\,\mathrm{d}\mathbf{x}_{k-1}, \tag{2}$$

where the transition pdf is given by $p(\mathbf{x}_k|\mathbf{x}_{k-1}) = p_{\mathbf{w}_k}(\mathbf{x}_k - \mathbf{F}_k\mathbf{x}_{k-1})$.
Correction (posterior update):

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) \propto p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1}), \tag{3}$$

where the likelihood is given by

$$p(\mathbf{y}_k|\mathbf{x}_k) = \prod_{j=1}^{n_{\mathbf{y}_k}} p_{v_{k,j}}(y_{k,j} - h_{k,j}(\mathbf{x}_k)). \tag{4}$$

The initial condition for the recursion is given by the pdf of the initial state $p(\mathbf{x}_0|\mathbf{y}_{1:0}) = p(\mathbf{x}_0)$. Knowledge of the posterior distribution (3) enables us to compute a state estimate that is optimal with respect to a given criterion. For example, the minimum mean-square error (MMSE) estimate is the posterior mean of $\mathbf{x}_k$ [10], [15]. In general and in the case analyzed within this paper, the conditional pdf cannot be determined analytically. The following section presents approximative methods for computing the posterior mean.

3 EKF and (Generalized) Gaussian Mixture Filter

In this section, we briefly describe the EKF. We then discuss GMF and GGMF, and quantify the complexity of the EKF and GGMF.

3.1 Extended Kalman Filter

The EKF, which applies Kalman filtering to a local linearization of the system (1), starts from a Gaussian approximation of the posterior at time t_{k-1}, with mean $\hat{\mathbf{x}}_{k-1}$ and covariance $\mathbf{P}_{k-1}$. The prediction (2) can then be solved exactly, leading to a Gaussian with mean $\hat{\mathbf{x}}_k^-$ and covariance $\mathbf{P}_k^-$, given by

$$\hat{\mathbf{x}}_k^- = \mathbf{F}_k\hat{\mathbf{x}}_{k-1},$$

$$\mathbf{P}_k^- = \mathbf{F}_k\mathbf{P}_{k-1}\mathbf{F}_k^T + \mathbf{Q}_k.$$

The correction step is based on a linearized measurement equation, leading to a Gaussian with posterior mean $\hat{\mathbf{x}}_k$ and posterior covariance matrix $\mathbf{P}_k$, given by

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{h}_k(\hat{\mathbf{x}}_k^-)\right)$$

$$\mathbf{P}_k = \mathbf{P}_k^- - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_k^-,$$

where $\mathbf{K}_k$ denotes the optimal Kalman gain

$$\mathbf{K}_k = \mathbf{P}_k^-\mathbf{H}_k\left(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{R}_k\right)^{-1}$$

1. It is straightforward to modify the paper's filters to deal with less restrictive assumptions but for the sake of simplicity we stay with these standard textbook assumptions, which are sufficient for our experimental setup.

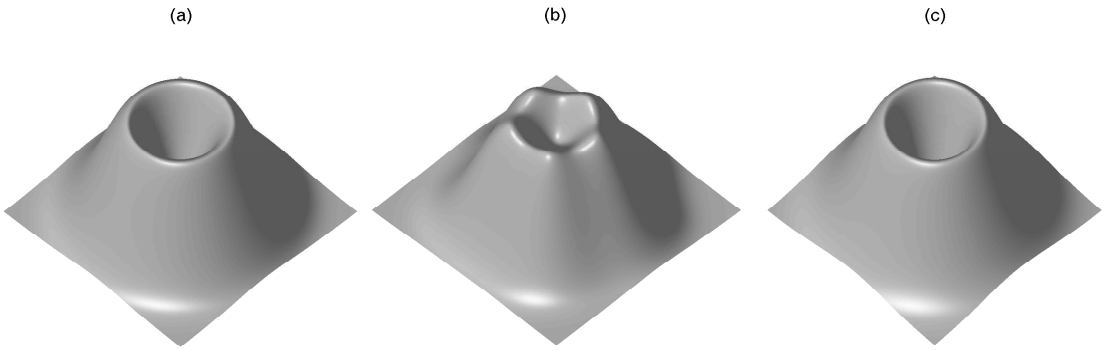


Fig. 1. Visualization of exact likelihood and approximations by GM and GGM: Subfigure (a) shows the exact normalized likelihood for a two-dimensional isotropic ranging model. The AP is in the center of the volcano-shaped likelihood and the "volcano's" brow corresponds to a circle with the range measurement as radius around the AP location. Subfigure (b) shows the approximation of the normalized likelihood by a GM with five components, and subfigure (c) shows the approximation of the normalized likelihood by a GGM with two components.

and $\mathbf{H}_k$ is obtained from the linearization of the measurement model around $\hat{\mathbf{x}}_k^-$:

$$\mathbf{H}_k = \left.\frac{\partial \mathbf{h}_k(\mathbf{x}_k)}{\partial \mathbf{x}_k}\right|_{\hat{\mathbf{x}}_k^-}.$$

It is known that in highly nonlinear situations the EKF can significantly underestimate the posterior covariance [13].

3.2 Gaussian Mixture Filter

In our context, the GMF also assumes a Gaussian state distribution at time t_{k-1}, with mean $\hat{\mathbf{x}}_{k-1}$ and covariance $\mathbf{P}_{k-1}$. Hence, the prediction step will be the same as with EKF, leading to a Gaussian distribution $\mathcal{N}(\mathbf{x}_k;\ \hat{\mathbf{x}}_k^-, \mathbf{P}_k^-)$.

The GMF then approximates the likelihood function (4) as follows: each of the $n_{\mathbf{y}_k}$ factors in (4) is approximated by a Gaussian mixture with N_j components ($j \in \{1, \ldots, n_{\mathbf{y}_k}\}$). The product of these mixtures is then a GM with $N = \prod_j N_j$ components. That GM is then multiplied with the prior, leading to a posterior with N components, with means $\boldsymbol{\mu}_i$, covariances $\boldsymbol{\Sigma}_i$, and weights $\lambda_i \in [0, 1]$. Before moving on to the next time step, this Gaussian mixture is collapsed to a single Gaussian. This can be done by moment matching, using the following formulas (derived in [14]) for the mean and covariance of the GM:

$$\mathrm{E}(\mathbf{x}) = \sum_{i=1}^{N} \lambda_i \boldsymbol{\mu}_i \triangleq \boldsymbol{\mu} \tag{5}$$

$$\mathrm{V}(\mathbf{x}) = \sum_{i=1}^{N} \lambda_i \left(\boldsymbol{\Sigma}_i + (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T\right). \tag{6}$$

Clearly, the complexity of this approach is prohibitive when $n_{\mathbf{y}_k}$ or N_j are large.

Fig. 1 (b) visualizes the principle of GMs for a two-dimensional isotropic Gaussian ranging model, which will be considered in Sections 4 and 5. In order to approximate the normalized likelihood $p(y_{k,j}|\mathbf{x}_k)$ in Fig. 1 (a) we have to pick N_j peaks of the exact likelihood and use them as mean values for the GM's N_j Gaussian components. Since the exact likelihood has an infinite number of peaks (the "volcano" brow in Fig. 1 (a)) we need to find a tradeoff between approximation quality and required computational resources.

Comment: Instead of collapsing to a single Gaussian after each correction step, more complex methods can be considered. Possible methods for reduction include forgetting and merging [16], [17], [18], [19], Gaussian Mixture Reduction via Clustering (GMRC) [20], and iterative compression algorithms [21]. For a broader overview we refer the reader to [22] and references therein. However, those reduction methods generally are only suitable for Gaussian mixtures with non-negative component weights.

3.3 Generalized Gaussian Mixture Filter

The GGMF addresses the complexity issue of the GMF by approximating likelihood functions associated with ranging measurements using a normalized mixture of only two Gaussians, but allowing *negative* weights. Thus, the likelihood of ranging measurement $y_{k,j}$ is approximated as

$$\begin{aligned} p(y_{k,j}|\mathbf{x}_k) \quad &\approx \quad \mathcal{N}(\mathbf{m}_1(y_{k,j});\ \boldsymbol{\mu}_{k,j}^{(1)}, \boldsymbol{\Sigma}_{k,j}^{(1)}) \\ &\quad \cdot \left(1 - \bar{c} \cdot \mathcal{N}(\mathbf{m}_2(y_{k,j});\ \boldsymbol{\mu}_{k,j}^{(2)}, \boldsymbol{\Sigma}_{k,j}^{(2)})\right) \end{aligned} \tag{7}$$

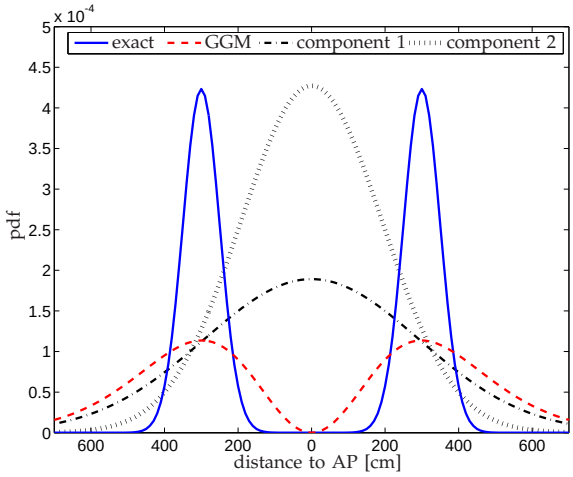


Fig. 2. Basic idea of GGM: Approximate a normalized likelihood (solid line) by a difference of two Gaussians (dashed line). Component 1 (dash-dotted line) gets positive weight, component 2 (dotted line) negative weight. The figure displays the projection of the normalized likelihood for an two-dimensional isotropic ranging model along its radial direction.

where $\bar{c} = c \cdot (2\pi)^{\frac{n_2}{2}} \sqrt{\det(\mathbf{\Sigma}_{k,j}^{(2)})}$, $c \leq 1$, and $\mathbf{m}_1(y_{k,j})$ and $\mathbf{m}_2(y_{k,j})$ are known functions of the measurements. If $\bar{c} = 0$ and $\mathbf{m}_1(y_{k,j}) = y_{k,j}$ then the conventional Kalman Filter computes the analytic solution. Using the approximate likelihood (7) yields a Gaussian mixture with possibly negative weights. However, all weights still have to sum up to one when the GM is normalized. Furthermore, parameter $\bar{c}$ ensures that such a likelihood is always nonnegative and therefore is a valid likelihood.

Fig. 1 (c) and Fig. 2 visualize the GGM's principle for an two-dimensional isotropic Gaussian ranging model. Instead of using some of the normalized likelihood's peaks (the "volcano" hilltop enclosure in Fig. 1 (c)), the GGM uses the "volcano" center, i.e. the AP location, as mean value for its two Gaussian components. More insights on how to use GGM for isotropic Gaussian ranging models will be presented in Section 4.1. The advantage of GGM compared with GM for the considered likelihood can be seen in Fig. 1. While the GGM approximation is smoother than the GM approximation and therefore resembles the likelihood more closely it requires significantly fewer components than the GM. Fig. 2 shows the projection along the radial direction of the normalized likelihood $p(y_{k,j}|\mathbf{x}_k)$ approximated by (7). Although (7) is only a rough approximation of the likelihood, it captures the location of the pdf's peaks and its general form much better than a single Gaussian (as in the EKF) or the GM (for a two-dimensional or higher-dimensional pdf) would.

TABLE 1
Complexity comparison: Number of operations from classes addition, subtraction, multiplication, division, and other (maximum, minimum, square, square root), dependent on number of measurements $n_{\mathbf{y}_k}$.

Filter	add	sub	mult	div	other
EKF	$\mathcal{O}(n_{\mathbf{y}_k}^2)$	$\mathcal{O}(n_{\mathbf{y}_k})$	$\mathcal{O}(n_{\mathbf{y}_k}^2)$	$\mathcal{O}(n_{\mathbf{y}_k})$	$\mathcal{O}(n_{\mathbf{y}_k})$
GGMF	$\mathcal{O}(2^{n_{\mathbf{y}_k}})$	$\mathcal{O}(2^{n_{\mathbf{y}_k}})$	$\mathcal{O}(2^{n_{\mathbf{y}_k}})$	$\mathcal{O}(2^{n_{\mathbf{y}_k}})$	$\mathcal{O}(2^{n_{\mathbf{y}_k}})$
EKF[a]	340	28	424	4	24
GGMF[a]	4,556	1,022	6,316	34	80

[a] for $n_{\mathbf{y}_k} = 4$ with four-dimensional $\mathbf{x}_k$

3.4 Computational complexity

It is clear that GGMF has a complexity that is exponential in the number of measurements $n_{\mathbf{y}_k}$. However, the complexity is much reduced compared to GMF for the problem considered within this paper, and still manageable for $n_{\mathbf{y}_k} < 5$, which is a reasonable value for practical scenarios (observe that in 3D, $n_{\mathbf{y}_k} = 4$ reference nodes suffice to obtain an unambiguous position estimate). A more detailed complexity comparison between GGMF and EKF is provided in Table 1 for addition, subtraction, multiplication, division, and other operations (maximum, minimum, square and square root).

4 Parameter Selection for GGMF

In this section, we describe how to select $\mathbf{m}_1(y_{k,j})$, $\mathbf{m}_2(y_{k,j})$, $\boldsymbol{\mu}_{k,j}^{(1)}$, $\boldsymbol{\mu}_{k,j}^{(2)}$, $\mathbf{\Sigma}_{k,j}^{(1)}$ and $\mathbf{\Sigma}_{k,j}^{(2)}$ in (7) for a positioning problem with Gaussian ranging errors. Then, we describe under which conditions GGM forms a good approximation.

4.1 Ranging Models

We consider isotropic Gaussian ranging models, which are frequently used in the literature (e.g. [23]), because the antennas of our UWB radios [24] used for all experiments were omnidirectional. The ranging models may be conditioned on specific propagation environments (such as line-of-sight (LOS) or non-line-of-sight (NLOS) propagation), which we will assume to be known. Moreover, we assume the mean of the ranging error is known a priori and that the ranging error variance σ^2 does not depend on the distance for ranges up to 50 meters. All these assumptions are based on experimental results with off-the-shelf UWB radios, and will be substantiated in Section 5.1. Hence, the likelihood function associated with a ranging measurement is given by

$$p(y_{k,j}|\mathbf{x}_k) \propto \exp\left(-\frac{1}{2\sigma^2}(y_{k,j} - ||\mathbf{x}_u^{(k)} - \mathbf{x}_{\mathrm{AP}}||)^2\right), \quad (8)$$

with $\mathbf{x}_u^{(k)}$ being the position vector contained in state $\mathbf{x}_k$. We will approximate this likelihood function with

a GGM centered at $\mathbf{x}_{\mathrm{AP}}$, meaning that $\boldsymbol{\mu}_{k,j}^{(1)} = \mathbf{x}_{\mathrm{AP}}$ and $\boldsymbol{\mu}_{k,j}^{(2)} = \mathbf{x}_{\mathrm{AP}}$, and $\boldsymbol{\Sigma}_{k,j}^{(1)} = \sigma_{\max}^2\mathbf{I}$ and $\boldsymbol{\Sigma}_{k,j}^{(2)} = \sigma_{\min}^2\mathbf{I}$ with $\sigma_{\min} < \sigma_{\max}$. Furthermore, $\mathbf{m}_1(y_{k,j}) = \mathbf{x}_u^{(k)}$ and $\mathbf{m}_2(y_{k,j}) = \mathbf{x}_u^{(k)}$. It remains to set $\sigma_{\min}$ and $\sigma_{\max}$.

We use a modified version of the approach used in [14] and model $\sigma_{\min}$ and $\sigma_{\max}$ as

$$\sigma_{\min} = \max\{\epsilon, \alpha y_{k,j} - \sigma\}, \tag{9}$$

$$\sigma_{\max} = \alpha y_{k,j} + \sigma, \tag{10}$$

where α is a configuration parameter that enables us to improve the approximation quality of our (normalized) GGM likelihood compared with the (normalized) exact likelihood. Thus, α can now be optimized with respect to a criterion (see Section 4.2 for an example). The value of ϵ ($\epsilon > 0$) in (9) ensures that the algorithm works also in cases where $\alpha y_{k,j} - \sigma \leq 0$. For the following analyses and tests we used $\epsilon = 0.1\,\mathrm{cm}$.

Finally, in order to ensure nonnegativity of the likelihood, we use $c = 1$. Therefore, under the Gaussian ranging error model, the only parameter remaining to be selected is the configuration parameter α.

4.2 Dependence of Approximation Quality on Ranging Variance

From Fig. 2 and the definition of the Gaussian distribution we expect that the approximation quality of a GGM depends on the standard deviation of the ranging errors. In order to quantify this dependence, we optimized α for a wide range of $\sigma \in [1, 200]\,\mathrm{cm}$. Optimization was performed so as to minimize the Kullback-Leibler divergence (KLD) [25] between the normalized GGM likelihood and the normalized exact likelihood. Recall that the KLD between a PDF $p(\cdot)$ and a PDF $q(\cdot)$ is defined as

$$D_{\mathrm{KL}}(p||q) = \int \ln\left(\frac{p(\mathbf{x})}{q(\mathbf{x})}\right) p(\mathbf{x})\,\mathrm{d}\mathbf{x}. \tag{11}$$

The integral (11) is evaluated numerically. Fig. 3 displays values of the KLD using α optimized by 1D line search at a range of ten meters. The results support our conjecture: as σ is increased, the KLD is reduced, meaning the GGM PDF approximates the exact PDF more accurately. For $\sigma \to 0$, the KLD blows up. Hence, for highly accurate ranging, GMM is ill-suited to approximate the exact PDF sufficiently well. In such cases it fails to model the steep probability gradients in the neighborhood of the range.

5 Simulations and Experiments

In order to get a better understanding under which circumstances the GGMF provides satisfying accuracy, we performed simulations, with models derived from experimental data, and compared the results with those of the EKF[2].

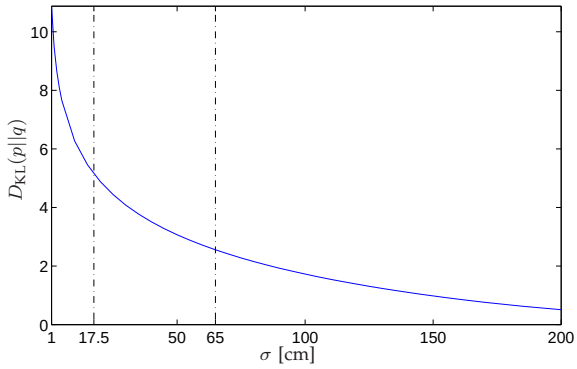


Fig. 3. Kullback-Leibler divergence (KLD) for different values of the ranging standard deviation σ using α optimized for a range of ten meters.

5.1 UWB Ranging Error Data and GGM Model

We carried out our experiments with a network of UWB radios. The radios (APs) used in our tests were Time Domain's PulsON 400 Ranging and Communications Modules (P400 RCM). They emit RF transmission from 3.1 GHz to 5.3 GHZ, with center frequency at 4.3 GHz, and provided us two-way time-of-flight (TW-TOF) ranging. Various studies have shown that in UWB networks TOA measurement noises, and thus TW-TOF measurement noises in LOS cases are usually very small [23], [26]. It has also been observed that NLOS measurements, in general, display significantly larger variances than LOS measurements [27], [28]. Therefore, we should trust LOS measurements more than NLOS measurements.

We collected 2 225 range measurements between ten APs placed in the known locations shown in Fig. 6, at different heights and different distances to each other. The used UWB radios have a built-in NLOS detection algorithm from which we labeled those range estimates as either LOS (1 745 measurements) or NLOS (480 measurements). Note that the equipment mislabeled some measurements as LOS and mislabeled some as NLOS. From this measurement data, we determined the corresponding ranging error standard deviations, σ_{LOS} and σ_{NLOS}. Fig. 4 shows the quantile-quantile (QQ) normal

2. We also implemented a GMF and a particle filter in MATLAB. However, primarily tests showed that it would be impractical for real-time positioning on small mobile devices since the GMF requires too many components for the likelihood approximations, and the particle filter requires too many particles. For example, with just 100 particles the computation time of the particle filter was more than 60 times higher than the EKF's and approximately 8 times higher than the GGMF's computation time for the scenarios 1, 3 and 5 described in Subection 5.2

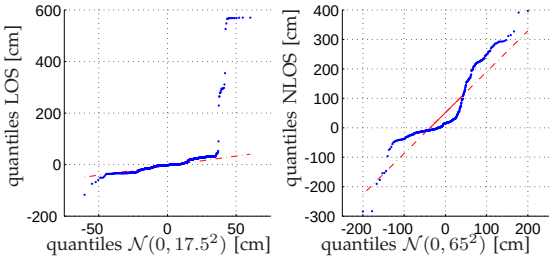


Fig. 4. QQ normal plots for errors in range estimates: The left plot shows the quantiles of ranging errors based on $1\,745$ LOS measurements versus the quantiles of a zero-mean Gaussian distribution with standard deviation $\sigma_{\text{LOS}} = 17.5\,\text{cm}$. The right plot shows the the quantiles of ranging errors based on 480 NLOS measurements versus the quantiles of a zero-mean Gaussian distribution with standard deviation $\sigma_{\text{NLOS}} = 65\,\text{cm}$.

plots that compare the quantiles of our LOS measurements (left figure) and NLOS measurements (right figure) with the quantiles of a zero-mean Gaussian distribution. Apart from a few outliers, the errors in LOS range estimates resemble such a Gaussian well. Since 68.26% of the absolute errors are smaller than $17.5\,\text{cm}$, we chose a zero-mean Gaussian distribution with standard deviation $\sigma_{\text{LOS}} = 17.5\,\text{cm}$ for modeling LOS ranging errors. In the right of the figure the QQ normal plot for the ranging errors supports the theory that NLOS measurements contain positive biases [1], and earlier findings that measurements made in NLOS conditions contain generally more outliers than measurements made in LOS condition [23]. We found that 68.44% of the absolute NLOS errors are smaller than $65\,\text{cm}$, and thus set $\sigma_{\text{NLOS}} = 65\,\text{cm}$.

Based on these values for σ_{LOS} and σ_{NLOS}, numerical minimization of the KLD at a nominal range of ten meters yielded the following optimal settings for the α parameter: $\alpha_{\text{LOS}} = 0.7374$ and $\alpha_{\text{NLOS}} = 0.7303$.

5.2 Simulation Setup

For our simulations the state

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_u \\ \mathbf{v}_u \end{bmatrix} \tag{12}$$

consisted of user position vector $\mathbf{x}_u$ and user velocity vector $\mathbf{v}_u$ in an East-North (2D) or East-North-Up (3D) coordinate system. The motion model matrix was

$$\mathbf{F}_k = \begin{bmatrix} \mathbf{I} & \Delta t\mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \tag{13}$$

i.e. a constant velocity model, and the state transition noise $p_{\mathbf{w}_k} = \mathcal{N}(\mathbf{w}_k; \mathbf{0}, \mathbf{Q})$ with

$$\mathbf{Q} = 4^2 \begin{bmatrix} \frac{(\Delta t)^3}{3}\mathbf{I} & \frac{(\Delta t)^2}{2}\mathbf{I} \\ \frac{(\Delta t)^2}{2}\mathbf{I} & \Delta t\mathbf{I} \end{bmatrix}. \tag{14}$$

Within the simulations we used the time step $\Delta t = 1$ second.

In total we simulated for six different scenarios 100 tracks of 100 time steps each: scenarios 1, 3, and 5 were in 2D, where we randomly distributed four APs uniformly in a $20\,\text{m}$ by $20\,\text{m}$ square; scenarios 2, 4, and 6 were in 3D, where we distributed the four APs uniformly in a $20\,\text{m}$ by $20\,\text{m}$ by $3.5\,\text{m}$ cube. At time step $t_k = 1$ those APs had either LOS or NLOS connection to the target node on the simulated track. The range estimates were generated according to the model from Section 5.1. The LOS/NLOS condition was simulated according to a Markov jump process: at each time step, the probability of staying in a LOS condition is p_{00}, while the probability of moving from a NLOS to a LOS condition is $1 - p_{11}$. The LOS/NLOS condition of node j at time k is denoted $\beta_{k,j} \in \{\text{LOS}, \text{NLOS}\}$.

For the six different scenarios initial values for the indicator variables and transition probabilities were chosen as follows:

- Scenario 1: 2D, $\beta_{1,j} = \text{LOS}$ for $j \in \{1, 2, 3, 4\}$, $p_{00} = p_{11} = 1$;
- Scenario 2: 3D, $\beta_{1,j} = \text{LOS}$ for $j \in \{1, 2, 3, 4\}$, $p_{00} = p_{11} = 1$;
- Scenario 3: 2D, $\beta_{1,j} = \text{LOS}$ for $j \in \{1, 2\}$ and $\beta_{1,j} = \text{NLOS}$ for $j \in \{3, 4\}$, $p_{00} = p_{11} = 0.8$;
- Scenario 4: 3D, $\beta_{1,j} = \text{LOS}$ for $j \in \{1, 2\}$ and $\beta_{1,j} = \text{NLOS}$ for $j \in \{3, 4\}$, $p_{00} = p_{11} = 0.8$;
- Scenario 5: 2D, $\beta_{1,j} = \text{NLOS}$ for $j \in \{1, 2, 3, 4\}$, $p_{00} = p_{11} = 1$; and
- Scenario 6: 3D, $\beta_{1,j} = \text{NLOS}$ for $j \in \{1, 2, 3, 4\}$, $p_{00} = p_{11} = 1$.

Scenarios 1 and 2 simulate benign cases with full LOS conditions, whereas scenarios 3 and 4 simulate more realistic cases, where both LOS and NLOS measurements occur. Scenarios 5 and 6, finally, simulate harsh conditions in which we can rely only on uncertain NLOS measurements. For all tracks and all scenarios the initial position estimate for EKF and GGMF was chosen at the center of the square (2D) or cube (3D) respectively with the initial covariance matrix set to an uninformative $10^6\mathbf{I}$. Furthermore, in the GGMF the posterior GGM was approximated by one Gaussian using (5) and (6) to ensure reasonable computation time. The mean of this Gaussian was used as position estimate.

We computed in each scenario for each track and each time step the positioning error as Euclidean distance, namely

$$e_{\hat{\mathbf{x}}_u} = ||\hat{\mathbf{x}}_u - \mathbf{x}_u||_2, \tag{15}$$

where $\hat{\mathbf{x}}_u$ and $\mathbf{x}_u$ are the estimated and the true user position. For evaluation we used the following four accuracy measures:

- *Mean error:* empirical mean of all 100-by-100 two- or three-dimensional positioning errors;

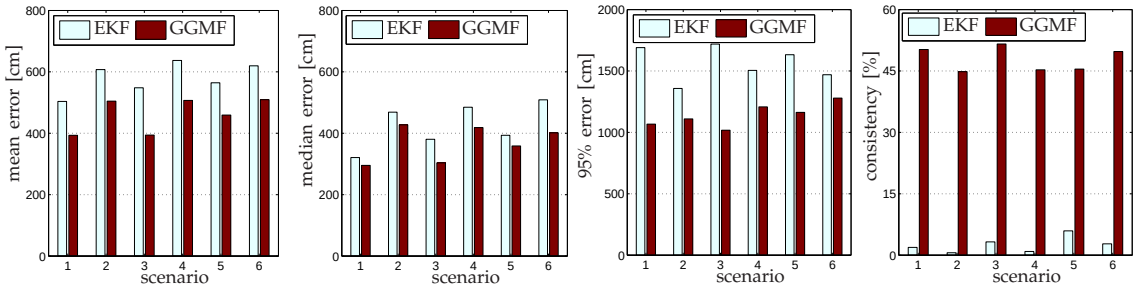


Fig. 5. Accuracy measures for simulation scenarios 1 to 6: In scenarios 1 (2D) and 2 (3D) four LOS measurements were available at every time step. In scenarios 3 (2D) and 4 (3D) at time $t_k = 1$ two LOS and two NLOS measurements were available, and their type could change at every time step according to a Markov jump process. In scenarios 5 (2D) and 6 (3D) four NLOS measurements were available at every time step.

- *Median error:* empirical median of all 100-by-100 two- or three-dimensional positioning errors;
- *95% error:* 95th percentile of all 100-by-100 two- or three-dimensional positioning errors; and
- *Consistency:* measures accuracy of the estimated position's covariance matrix; percentage of time steps for which a filter was consistent with respect to the Gaussian consistency test [10, p. 235 ff.] with risk level 5%.[3]

Comment: In our simulations neither EKF nor GGMF tested if their current position estimate was feasible (e.g., within the area in which the tracks were simulated). Using map information could significantly improve results, but is out of the scope for this paper.

5.3 Simulation Results

Fig. 5 contains the accuracy measures mean, median, 95% errors and consistency levels for all six simulation scenarios for both EKF and GGMF.

The GGMF outperforms the EKF in all six scenarios. Mean errors errors decreased 16% to 28%, median errors decreased between 8% and 21%, and 95% errors decreased 13% to 41%. Moreover, the GGMF provided significantly higher consistency levels than the EKF. However, both filters show consistency far below the desired 95%. This is expected for the EKF, since it tends to underestimate the true covariance matrix due to nonlinearities [13]. The weak consistencies for our GGMF might be a result of using the Gaussian consistency test that assumes Gaussian distributions, which does not hold in the analyzed cases. Alternatively, we could have used the general inconsistency test [13], which should report higher consistency levels for all tests with all filters.

3. In the Gaussian consistency test, for a risk level of 5%, a filter is said to be consistent at a certain time step if the estimated covariance matrix $\mathbf{P}_{k,\hat{\mathbf{x}}_u}$ of the estimated user position $\hat{\mathbf{x}}_u$ fulfills the inequality $(\hat{\mathbf{x}}_u - \mathbf{x}_u)^T \mathbf{P}^{-1}_{k,\hat{\mathbf{x}}_u} (\hat{\mathbf{x}}_u - \mathbf{x}_u) \leq \chi^2_2(0.95) = 5.9915$, where $\mathbf{x}_u$ is the true user position. In case of Gaussian posterior distribution, the closer the consistency value is to 95%, the better the covariance matrix estimation is.

We also determined the computation time of EKF and GGMF for our particular implementation. In line with our results from Section 3.4, the computation time for GGMF for all six scenarios is approximately eight times higher than for EKF.

Furthermore, we applied an unscented Kalman filter (UKF) on scenarios 1, 3 and 5 [29]. The obtained accuracy measures were very close to those of the EKF, except for better consistency (but still worse than GGMF's consistency). Furthermore, the UKF needed approximately three times more computation time than the EKF. Thus, for better readability of the figures UKF results are not shown.

5.4 Experimental Results

We now report our evaluation of EKF and GGMF with real-world data. We placed ten UWB radios as reference nodes in a gym at Chalmers University of Technology. An additional radio was carried by a person, attached to a belt at hip height, and moved along a track denoted by "ground truth", shown in Fig. 6, to collect ranging measurements. During the measurement campaign not all APs provided range estimates at each point of the track.[4] Furthermore, during our measurement campaign there were several gymnastics apparatuses in the gym, which are not marked in the map. We carried out four distinct tests:

- Test 1: two-dimensional positioning, using all reference nodes;
- Test 2: three-dimensional positioning, using all reference nodes;
- Test 3: two-dimensional positioning, using only 5 reference nodes;
- Test 4: three-dimensional positioning, using only 5 reference nodes.

For each test, we evaluated the performance in terms of our accuracy measures, shown in Table 2.

4. The UWB radios were powered by batteries, which for one radio failed to provide power for the whole test track. Since such failures can also occur in real-world applications, we decided to use this AP as long as it was functioning.

TABLE 2
Accuracy measures for tests with real data. Column *time* gives the relative computation time using a specific MATLAB implementation, scaled up in each test so that computation time for EKF is 1.

test	filter	time	mean [cm]	median [cm]	95% err. [cm]	cons.
1	EKF	1	137	87	336	4 %
1	GGMF	81	88	90	133	100 %
2	EKF	1	303	93	2 199	6 %
2	GGMF	70	99	95	160	100 %
3	EKF	1	168	75	465	10 %
3	GGMF	5	111	108	204	91 %
4	EKF	1	1 155	190	4 453	0 %
4	GGMF	5	137	135	219	88 %

In all cases, the GGMF clearly outperformed the EKF. Its consistency levels, mean and 95% errors are significantly better. However, for tests 1 to 3 the EKF yields similar or smaller median errors. Fig. 6 and 7 show how the errors and error statistics evolved for the two-dimensional positioning. Whereas the GGMF provides satisfying position estimates from the first time step on, the EKF needs several steps until it provides position estimates close to the true positions (we observed a similar behavior in our simulations). Once it yields an estimate close to the true position the EKF performs very well; even better than the GGMF. This can be explained by the fact that due to nonlinearities the EKF underestimates the true covariance matrix [13], and thus trusts its prior more than the GGMF. Furthermore, once the EKF yields a satisfying positioning estimate the approximation quality of the measurement function improves (since the linearization is performed in the prior mean), which improves the filter's positioning estimates.

Fig. 8 shows for tests 1 and 3 the square root of the trace of the posterior covariance matrices $\mathbf{P}_{k,\hat{\mathbf{x}}_u}$, which is associated to the estimated user position $\hat{\mathbf{x}}_u$. For the GGMF those values are generally above or close to the errors of the GGMF's position estimates. In tests 1 and 2 the algorithm overestimates the posterior covariance, whereas it underestimates it when using only five reference nodes. By contrast the EKF's position errors are generally significantly larger than $\|\mathbf{P}_{k,\hat{\mathbf{x}}_u}\|_2 = \sqrt{\mathrm{tr}(\mathbf{P}_{k,\hat{\mathbf{x}}_u})}$, which results in very poor consistencies for the tests and supports the earlier finding that the EKF tends to underestimate the true covariance matrix [13].

6 Conclusion

We have studied the generalized Gaussian mixture filter (GGMF), a Bayesian low-complexity localization and navigation algorithm, which can deal with significant nonlinearities. GGMF approximates the likelihood associated with range measurements in an isotropic Gaussian ranging model as a mixture of two Gaussian components, one having negative weight. For Gaussian ranging errors, we showed that the approximation quality of GGMF is improved for larger ranging error variances. Due to this, the GGMF is able to outperform the standard extended Kalman filter (EKF) in situations where large ranging errors are possible and where network geometry is poor. For scenarios with small ranging error variances the 2-component GGM captures the shape of the isotropic ranging error model only roughly, wherefore the GGMF offers only similar positioning accuracy than the EKF. Because the GGMF was originally developed for positioning using highly uncertain cellular telephone measurements, this behavior was expected. In addition, the tests showed that applying the GGMF is advantageous particularly in the starting phase of the positioning. Once a sufficiently precise position estimate is yielded, also the EKF provides satisfying position estimates, given a reasonable motion model. In all cases, the GGMF comes with an increase in computational complexity, which would allow real-time positioning on small mobile devices as long as the number of used range estimates at each time step is kept low (e.g., below 5). Our findings are corroborated with extensive simulations (based on real UWB ranging data) and experimental results.

Future work includes the extension to WLAN and Bluetooth localization, and the comparison with methods tailored to highly nonlinear conditions (e.g., fingerprinting, particle filtering). We will also

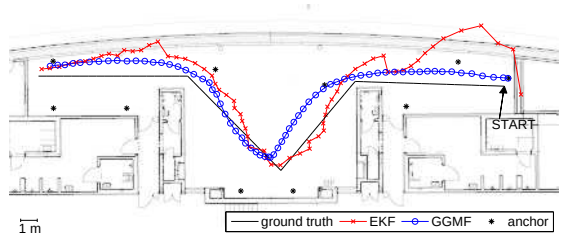


Fig. 6. Positioning in 2D using ten reference nodes (test 1).

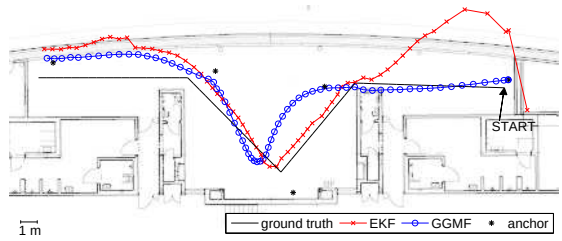


Fig. 7. Positioning in 2D using five reference nodes (test 3).

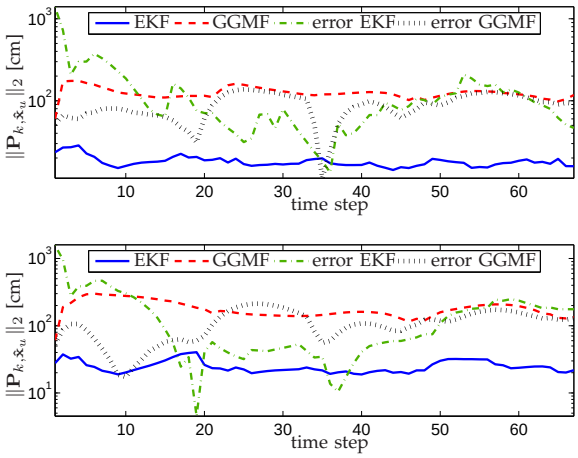


Fig. 8. $\|\mathbf{P}_{k,\hat{\mathbf{x}}_u}\|_2 = \sqrt{\text{tr}(\mathbf{P}_{k,\hat{\mathbf{x}}_u})}$ for time steps t_1 to t_{67} for test 1 (upper plot) and test 3 (lower plot) compared with positioning errors for EKF and GGMF.

investigate how GGMF can be adapted to scenarios with very small ranging error variances and other system models than the model considered within this paper, and how the complexity can be further reduced. One possible approach for handling highly reliable range measurements, which will be studied and compared with particle filters and geometric techniques in the future, is the use of other distributions, for example Student-t distributions, for the mixtures.

Acknowledgments

We would like to thank Matti Raitoharju, Henri Nurminen, Simo Ali-Löytty, Juha Ala-Luhtala, and Yakup Kiliç for their careful reading of the manuscript and valuable suggestions. We are also grateful to Yakup Kiliç, Pinar Oguz Ekim, Gabriel E. Garcia, and Christopher Lindberg for their involvement in the UWB measurement campaign.

References

[1] S. Gezici, Z. Tian, G. B. Biannakis, H. Kobayashi, A. F. Molisch, H. V. Poor, and Z. Sahinoglu, "Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 70–84, July 2005.

[2] F. Gustafsson and F. Gunnarsson, "Mobile positioning using wireless networks: possibilities and fundamental limitations based on available wireless network measurements," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 41–53, 2005.

[3] E. D. Kaplan, *Understanding GPS: Principles and Applications*, 2nd ed., E. D. Kaplan and C. J. Hegarty, Eds. Artech House, November 2005.

[4] D. B. Jourdan, D. Dardari, and M. Z. Win, "Position error bound for UWB localization in dense cluttered environments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 2, pp. 613–628, April 2008.

[5] B. Alavi and K. Pahlavan, "Modeling of the TOA-based distance measurement error using UWB indoor radio measurements," *IEEE Communications Letters*, vol. 10, no. 4, pp. 275–277, April 2006.

[6] D. Dardari, A. Conti, U. Ferner, A. Giorgetti, and M. Z. Win, "Ranging with ultrawide bandwidth signals in multipath environments," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 404–426, February 2009.

[7] P. S. Maybeck, *Stochastic Models, Estimation, and Control Volume 1*, ser. Mathematics in Science and Engineering. Academic Press, 1979, vol. 141.

[8] S. J. Julier and J. K. Uhlmann, "A new extension of the Kalman filter to nonlinear systems," in *Proceedings of AeroSense: the 11th international symposium on aerospace/defence sensing, simulation and controls*, 1997.

[9] L. Mihaylova, D. Angelova, S. Honary, D. Bull, C. Canagarajah, and B. Ristic, "Mobility tracking in cellular networks using particle filtering," *IEEE Transactions on Wireless Communications*, vol. 6, no. 10, pp. 3589–3599, 2007.

[10] Y. Bar-Shalom, R. X. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation, Theory Algorithms and Software*. John Wiley & Sons, 2001.

[11] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, ser. Mathematics in Science and Engineering. Academic Press, 1970, vol. 64.

[12] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[13] S. Ali-Löytty, N. Sirola, and R. Piché, "Consistency of three Kalman filter extensions in hybrid navigation," in *Proceedings of The European Navigation Conference GNSS 2005*, Munich, Germany, Jul. 2005.

[14] P. Müller, S. Ali-Löytty, M. Dashti, H. Nurminen, and R. Piché, "Gaussian mixture filter allowing negative weights and its application to positioning using signal strength measurements," in *9th Workshop on Positioning Navigation and Communication (WPNC)*, 2012, pp. 71–76.

[15] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter, Particle Filters for Tracking Applications*. Boston, London: Artech House, 2004.

[16] S. Ali-Löytty and N. Sirola, "Gaussian mixture filter in hybrid navigation," in *Proceedings of The European Navigation Conference GNSS 2007*, May 2007, pp. 831–837.

[17] S. Ali-Löytty, "Gaussian mixture filters in hybrid positioning," Ph.D. dissertation, Tampere University of Technology, August 2009. [Online]. Available: http://URN.fi/URN:NBN:fi:tty-200905191055

[18] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, no. 4, pp. 465–479, July 1971.

[19] D. J. Salmond, "Mixture reduction algorithms for target tracking," *State Estimation in Aerospace and Tracking Applications, IEE Colloquium on*, pp. 7/1–7/4, 1989.

[20] D. Schieferdecker and M. F. Huber, "Gaussian mixture reduction via clustering," in *12th International Conference on Information Fusion (FUSION '09)*, 2009, pp. 1536–1543.

[21] D. F. Crouse, P. Willett, K. Pattipati, and L. Svensson, "A look at Gaussian mixture reduction algorithms," in *Proceedings of the 14th International Conference on Information Fusion (FUSION)*, 2011, pp. 1–8.

[22] M. Kristan, D. Skočaj, and A. Leonardis, "Incremental learning with Gaussian mixture models," in *Computer Vision Winter Workshop 2008, Janez Perš (ed.)*, 2008, pp. 25–32.

[23] H. Wymeersch, J. Lien, and M. Z. Win, "Cooperative localization in wireless networks," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 427–450, 2009.

[24] "P400 RCM data sheet," July 2011. [Online]. Available: http://www.timedomain.com/datasheets/320-0289BP400RCMDataSheetFinal.pdf

[25] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. New York, NY, 10158: John Wiley & Sons, 1991.

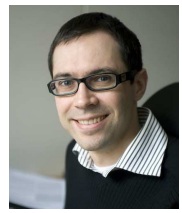
[26] Z. Low, J. Cheong, C. Law, W. Ng, and Y. Lee, "Pulse detection algorithm for line-of-sight (LOS) UWB ranging applications," *IEEE Antennas and Wireless Propagation Letters*, vol. 4, pp. 63–67, 2005.

[27] M. I. Silventoinen and T. Rantalainen, "Mobile station emergency locating in GSM," in *Proceedings of IEEE International Conference on Personal Wireless Communications*, 1996, pp. 232–238.

[28] M. P. Wylie and J. Holtzman, "The non-line of sight problem in mobile location estimation," in *Proc. 5th IEEE Int. Conf. Universal Personal Communications*, September 1996, pp. 827–831.

[29] E. Wan and R. V. der Merwe, "The unscented Kalman filter," in *Kalman filtering and neural networks*, S. Haykin, Ed. Wiley, 2001, ch. 7.

Philipp Müller received the M.Sc. degree in Mathematics in 2010 from the Chemnitz University of Technology, Germany. He is currently a Ph.D. student in the Department of Automation Science and Engineering at Tampere University of Technology, Finland. His research interests are algorithms for low-complexity localization.



Henk Wymeersch (S'99, M'05) received the Ph.D. degree in Electrical Engineering/Applied Sciences in 2005 from Ghent University, Belgium. He is currently an Associate Professor with the Department of Signals and Systems at Chalmers University of Technology, Sweden. He is also affiliated with the FORCE research center on fiber-optic communication, and is the PI of COOPNET an ERC project on cooperative networks. Prior to joining Chalmers, he was a Postdoctoral Associate with the Laboratory for Information and Decision Systems (LIDS) at the Massachusetts Institute of Technology (MIT). He is a member of the IEEE, served as Associate Editor for IEEE COMMUNICATION LETTERS (2009–2013) and the TRANSACTIONS ON EMERGING TELECOMMUNICATIONS TECHNOLOGIES (2011–present). He served as Guest Editor for EURASIP JOURNAL ON WIRELESS COMMUNICATIONS AND NETWORKING (special issue on Localization in Mobile Wireless and Sensor Networks), and for EURASIP JOURNAL ON ADVANCES IN SIGNAL PROCESSING (special Issue on Signal Processing Techniques for Anywhere, anytime positioning). He has co-authored over 100 contributions in journals and international conferences, and is the author of Iterative Receiver Design (Cambridge University Press, August 2007). His research interests include algorithm design for wireless transmission, statistical inference, and iterative processing.

Robert Piché (M'10) received the Ph.D. degree in Civil Engineering in 1986 from the University of Waterloo, Canada. He is Professor in the Department of Automation Science and Engineering at Tampere University of Technology, Finland. His scientific interests are in mathematical modelling, numerical analysis, and systems theory and in applications including positioning, finance, and solid mechanics.

PUBLICATION 4

Philipp Müller and Robert Piché: Statistical trilateration with skew-t errors. In *2015 International Conference on Localization and GNSS (ICL-GNSS)*, 1–6, Gothenburg, Sweden, June 2015.
DOI: 10.1109/ICL-GNSS.2015.7217164

Statistical Trilateration with Skew-t Errors

Philipp Müller and Robert Piché
Department of Automation Science and Engineering (ASE)
Tampere University of Technology
P.O. Box 692, FI-33101 Tampere, Finland
Emails: philipp.muller@tut.fi, robert.piche@tut.fi

***Abstract*—In the problem of determining a target's location using radio signal time-of-flight to reference nodes with known locations, measurement errors can be skewed because of multipath effects. In this paper, range errors are modelled using the skew-t distribution. An Expectation-Maximisation (EM) algorithm for computing the unknown location is presented, and its accuracy is compared with a descending Gauss-Newton algorithm by simulations. The EM algorithm improves the positioning accuracy significantly. Furthermore, it is shown how to fit the parameters of a skew-t distribution to training data using a Gibbs sampler.**

I. Introduction

Trilateration is the process of determining the position of a target using measurements of distances (ranges) to reference nodes with known positions. The target's position can be computed as the solution of a nonlinear least squares (NLS) problem that is equivalent to nonlinear regression with an elliptical (e.g. normal or t-distributed) error model. Computational algorithms for NLS problems are well known, and in trilateration problems they outperform "closed-form" methods [1].

A drawback of using elliptical distributions is that they cannot model skewness in the distribution of measurement errors since they are symmetric. Skewness can arise in time-of-flight range measurements because of multipath effects. For example, the non-line-of-sight data from UWB network in [2] is clearly skewed (Fig. 4). There is therefore interest in extending trilateration methods to noise models that include skewness. The skew-t distribution, which is a parametric distribution family that includes the t-distribution as a special case, is well-suited for this purpose. There is a considerable body of research related to this distribution family in the statistical literature, and an extensive discussion can be found, for example, in Azzalini's recent monograph [3]. Much of the literature focuses on fitting the parameters of a skew-t distribution using methods such as the Expectation-Maximisation (EM) algorithm to given error data.

The object of this paper is to show how to apply skew-t statistical theory and methods to the trilateration problem, where the parameters of the skew-t distribution are known or have already been fitted to training data. For positioning the EM algorithm is used, which is an attractive choice for nonlinear regression because its monotone convergence ensures numerical stability. In addition, in nonlinear regression one can use a standard NLS solver inside the EM iteration loop.

This paper is organised as follows. The trilateration problem and basic properties of the skew-t distribution are reviewed in Section II. Section III presents an EM algorithm for solving the trilateration problem under the assumption of skew-t distributed measurement errors. The algorithm's performance is tested in Section IV and it is compared with a descending Gauss-Newton method. Section V explains briefly how to fit the parameters to given error data using a Gibbs sampler, and presents an example for the fitting. Some concluding remarks and an outlook are given in Section VI.

Notation : $\mathbf{x}$ and $x_{1:d}$ denote column vectors, $\mathbf{H}$ denotes a matrix, and underscores are used to denote random variables and random vectors in contexts where the distinction from deterministic variables is useful.

II. Model

A. Statistical trilateration

We use the following statistical formulation of the trilateration problem [4].

Let the unknown target location be represented by the d-dimensional random vector $\underline{x}_{1:d} = \underline{\mathbf{x}}$. The K scalar measurements are modelled as

$$\underline{y}_k \,|\, (\underline{\mathbf{x}} = \mathbf{x}) = h_k(\mathbf{x}) + \underline{v}_k \tag{1}$$

for $k \in 1,..,K$, where function $h_k : \mathbb{R}^d \to \mathbb{R}$ models the measurement geometry and $\underline{v}_1, .., \underline{v}_K$ are mutually independent random variables (additive errors). In addition, $\underline{v}_{1:K}$ and $\underline{\mathbf{x}}$ are independent.

The prior probability density function (pdf) of $\mathbf{x}$ is denoted as $p_{\underline{\mathbf{x}}}$, and the pdf of v_k as $p_{\underline{v}_k}$. The posterior distribution of $\mathbf{x}$ given the K-dimensional measurement vector $\mathbf{y}_{1:K}$ has the pdf

$$p_{\underline{\mathbf{x}}|\underline{\mathbf{y}}_{1:K}}(\mathbf{x}|\mathbf{y}_{1:K}) \propto p_{\underline{\mathbf{x}}}(\mathbf{x}) \prod_{k=1}^{K} p_{\underline{v}_k}\left(y_k - h_k(\mathbf{x})\right) \tag{2}$$

A value of $\mathbf{x}$ that maximises (2) is called a maximum *a posteriori* (MAP) estimate. This MAP estimate coincides with the maximum likelihood (ML) estimate if the prior distribution is "flat", i.e. if $p_{\underline{\mathbf{x}}}(\mathbf{x}) \propto 1$.

In trilateration, the measurement function is the Euclidean distance between the target and a reference node at a known location $\mathbf{c}_k$:

$$h_k(\mathbf{x}) = ||\mathbf{x} - \mathbf{c}_k||. \tag{3}$$

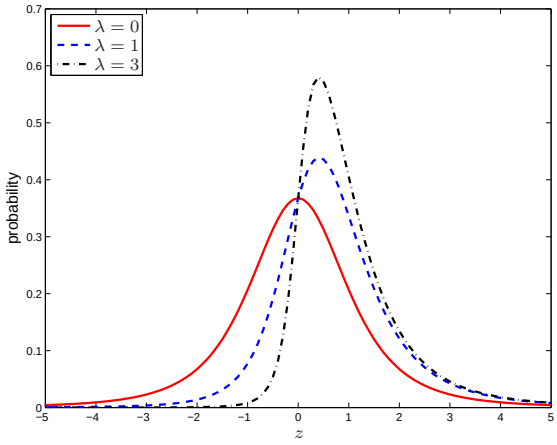


Fig. 1. Standardised skew-t distribution ($\nu = 3$).

The Jacobian of $\mathbf{h} = h_{1:K}$ is the $K \times d$ matrix $\mathbf{H}$ whose kth row is the transpose of a unit vector pointing from $\mathbf{c}_k$ to $\mathbf{x}$, that is,

$$\mathbf{H}_{k,1:d} = \nabla h_k(\mathbf{x}) = \frac{(\mathbf{x} - \mathbf{c_k})^T}{||\mathbf{x} - \mathbf{c}_k||}. \tag{4}$$

B. The skew-t distribution

In this subsection some general properties of the skew-t distribution are collected. For more extensive discussion see, for example, [3, pp. 101 ff.] and [5].

A random variable $\underline{z}$ is said to have a skew-t distribution with location ξ, scale σ^2, skewness λ and ν degrees-of-freedom (dof) if its pdf has the form

$$p_{\underline{z}}(z) = \frac{2}{\sigma} t_\nu \left(\frac{z-\xi}{\sigma}\right) T_{\nu+1}\left(\lambda \frac{z-\xi}{\sigma}\sqrt{\frac{\nu+1}{\nu + \frac{(z-\xi)^2}{\sigma^2}}}\right), \tag{5}$$

where t_ν and $T_{\nu+1}$ denote the pdf and the cumulative distribution function (cdf) of the standardised t-distribution. The skew-t distribution is denoted $\underline{z} \sim \mathrm{ST}\left(\xi, \sigma^2, \lambda, \nu\right)$.

The skew-t reduces to the t-distribution when $\lambda = 0$, that is, $\mathrm{ST}\left(\xi, \sigma^2, 0, \nu\right) = \mathrm{T}\left(\xi, \sigma^2, \nu\right)$, a t-distribution with location ξ, scale σ^2 and ν dof. The skew-t reduces to $\mathrm{N}\left(\xi, \sigma^2\right)$, a normal distribution with mean ξ and variance σ^2, when $\lambda = 0$ and $\nu \to \infty$. $\mathrm{ST}\left(0, 1, \lambda, \nu\right)$ is called a standardised skew-t distribution (Fig. 1) and if $\underline{u} \sim \mathrm{ST}\left(0, 1, \lambda, \nu\right)$ then $a + b\underline{z} \sim \mathrm{ST}\left(a, b^2, \lambda, \nu\right)$.

For $\nu > 2$ the mean and variance of $\underline{z} \sim \mathrm{ST}\left(\xi, \sigma^2, \lambda, \nu\right)$ are

$$\mathrm{E}(\underline{z}) = \xi + \sigma g_\nu \delta_\lambda \quad , \quad \mathrm{var}(\underline{z}) = \sigma^2\left(\frac{\nu}{\nu-2} - (g_\nu \delta_\lambda)^2\right) \tag{6}$$

where $g_\nu = \frac{\sqrt{\nu}\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}$ and $\delta_\lambda = \frac{\lambda}{\sqrt{1+\lambda^2}} \in (-1, 1)$.

The skew-t distribution has the following hierarchical model. Let $\underline{\tau} \sim \Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$ and $\underline{w} \sim \mathrm{N}\left(0, 1\right)$, which denote the Gamma distribution with both shape and scale $\frac{\nu}{2}$ and the standard normal distribution. Then $\underline{t} = |\frac{\sigma}{\sqrt{\underline{\tau}}}\underline{w}|$ is a half-normal (HN) random variable with pdf

$$p_{\underline{t}}(t) \propto \Phi\left(t\frac{\sqrt{\tau}}{\sigma}\right)[t > 0], \tag{7}$$

where $[.]$ is the Iverson bracket and Φ denotes the cdf of the standard normal distribution. Samples from the distribution $\mathrm{ST}\left(\xi, \sigma^2, \lambda, \nu\right)$ can be drawn from the conditional distribution

$$\underline{z}|(\underline{t} = t, \underline{\tau} = \tau) \sim \mathrm{N}\left(\xi + \frac{\lambda t}{\sqrt{1+\lambda^2}}, \frac{1-\delta_\lambda^2}{\tau}\sigma^2\right). \tag{8}$$

In the hierarchical representation (7 – 8), the conditional random variable $\underline{t}|(\underline{z} = z, \underline{\tau} = \tau)$ has the distribution $\mathrm{N}\left(\delta_\lambda(z-\xi), \frac{1-\delta_\lambda^2}{\tau}\sigma^2\right)$ and the conditional random variable $\underline{\tau}|(\underline{z} = z)$ has the pdf

$$\underline{\tau}|(\underline{z} = z) \propto \tau^{(\nu-1)/2}\exp\left(-\frac{\tau}{2}(\eta^2 + \nu)\right)\Phi(\lambda\eta\sqrt{\tau}), \tag{9}$$

where $\eta = \frac{z-\xi}{\sigma}$.

III. Estimating the target position

A. Positioning using Expectation Maximisation

This subsection presents an expectation maximisation (EM) algorithm to compute the MAP estimate, i.e. the mode of (2), for additive skew-t measurement errors. In the maximisation step (M-step) the descending Gauss-Newton (GN) algorithm is used. A more detailed derivation of the equations used in the algorithm can be found in Subsection III.B.

The posterior pdf (2) for the measurement model $\underline{y}_k|(\underline{\mathbf{x}} = \mathbf{x}) \sim \mathrm{ST}\left(\xi + h_k(\mathbf{x}), \sigma^2, \lambda, \nu\right)$ is

$$p_{\underline{\mathbf{x}}|\underline{\mathbf{y}}_{1:K}}(\mathbf{x}|\mathbf{y}_{1:K}) \propto p_{\underline{\mathbf{x}}}(\mathbf{x}) \prod_{k=1}^{K} \frac{2}{\sigma} t_\nu\left(\frac{\bar{y}_k}{\sigma}\right) T_{\nu+1}\left(\lambda\frac{\bar{y}_k}{\sigma}\sqrt{\frac{\nu+1}{\nu+\frac{\bar{y}_k^2}{\sigma^2}}}\right), \tag{10}$$

where $\bar{y}_k = y_k - h_k(\mathbf{x}) - \xi$. Using (8), a hierarchical version of the measurement model is

$$\underline{y}_k|(\underline{\mathbf{x}} = \mathbf{x}, \underline{t}_k = t_k, \underline{\tau}_k = \tau_k) \sim \mathrm{N}\left(\xi + h_k(\mathbf{x}) + \delta_\lambda t_k, \frac{1-\delta_\lambda^2}{\tau_k}\sigma^2\right), \tag{11}$$

where the hyperparameters are $\underline{t}_k = |\frac{\sigma}{\sqrt{\underline{\tau}_k}}\underline{w}_k|$ with $\underline{w}_k \sim \mathrm{N}\left(0, 1\right)$ and $\underline{\tau}_k \sim \Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$.

In the EM algorithm's expectation step (E-step) the hyperparameters are updated by setting them to the mean values of their conditional distribution, i.e.

$$\hat{\tau}_k \leftarrow \frac{\sigma^2\left(1-\delta_\lambda^2\right)\nu^2}{\left(y_k - h_k(\hat{\mathbf{x}}) - \xi - \delta_\lambda t_k\right)^2 + 4\sigma^2\left(1-\delta_\lambda^2\right)} \tag{12a}$$

$$\hat{t}_k \leftarrow \mu_k + \frac{\phi(-\mu_k/\sigma_k)}{1 - \Phi(-\mu_k/\sigma_k)}\sigma_k \tag{12b}$$

for $k = \{\, 1, .., K \,\}$, where $\hat{\mathbf{x}}$ is the current MAP estimate, and ϕ and Φ denote the pdf and the cdf of the standard normal

distribution. In (12b) $\mu_k = \frac{1}{2\delta_\lambda}\left(y_k - h_k(\hat{\mathbf{x}}) - \xi\right)$ and $\sigma_k = \sqrt{\frac{\sigma^2\left(1-\delta_\lambda^2\right)}{2\tau_k}}$.

Assuming a multivariate-normal prior distribution for $\mathbf{x}$ with mean $\mathbf{m}_0$ and covariance $\mathbf{P}_0$, the M-step is the optimisation of the conditional posterior

$$p(\mathbf{x}|\mathbf{y}_{1:K}, t_{1:K}, \tau_{1:K}) \propto \mathrm{e}^{-\frac{1}{2}\left((\mathbf{x}-\mathbf{m}_0)'\mathbf{P}_0^{-1}(\mathbf{x}-\mathbf{m}_0)+\frac{\sum_{k=1}^K \tau_k\left(\mathbf{y}_{1:K}-\mathbf{h}_{1:K}(\mathbf{x})-\xi-\delta_\lambda t_k\right)^2}{\sigma^2\left(1-\delta_\lambda^2\right)}\right)} \tag{13}$$

with $\tau_k = \hat{\tau}_k$ and $t_k = \hat{t}_k$, that is

$$\hat{\mathbf{x}} \leftarrow \arg\min_{\mathbf{x}} \left((\mathbf{x}-\mathbf{m}_0)'\mathbf{P}_0^{-1}(\mathbf{x}-\mathbf{m}_0) + \frac{\sum_{k=1}^K \hat{\tau}_k\left(y_k - h_k(\mathbf{x}) - \xi - \delta_\lambda \hat{t}_k\right)^2}{\sigma^2\left(1-\delta_\lambda^2\right)}\right) \tag{14}$$

The minimisation in (14) can be computed by any nonlinear least-squares optimisation method. In this paper the descending Gauss-Newton algorithm (see e.g. [4]) is used, which for (14) is the iteration of

$$\mathbf{H} \leftarrow \frac{\partial \mathbf{h}_{1:K}}{\partial \mathbf{x}}(\hat{\mathbf{x}}) \tag{15a}$$

$$\hat{\mathbf{x}} \leftarrow \hat{\mathbf{x}} + \alpha\left(\mathbf{m}_0 - \hat{\mathbf{x}} + \mathbf{K}\left(\tilde{\mathbf{y}} - \mathbf{h}(\hat{\mathbf{x}}) - \mathbf{H}\mathbf{m}_0 + \mathbf{H}\hat{\mathbf{x}}\right)\right) \tag{15b}$$

with $\mathbf{K} = \mathbf{P}_0\mathbf{H}^T\left(\mathbf{R} + \mathbf{H}\mathbf{P}_0\mathbf{H}^T\right)^{-1}$, where $\mathbf{R} = \sigma^2(1-\delta_\lambda^2)\mathrm{diag}\{1/\hat{\tau}_1, .., 1/\hat{\tau}_K\}$, and modified "data" $\tilde{\mathbf{y}} = \mathbf{y}_{1:K} - \xi - \delta_\lambda \hat{t}_{1:K}$. The scale factor α ensures that the cost function is decreasing, and is found by line search (see Subsection III.B).

In case of a flat prior, (15b) is replaced by

$$\hat{\mathbf{x}} \leftarrow \hat{\mathbf{x}} - \alpha\left(\left(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}\left(\mathbf{h}(\hat{\mathbf{x}}) - \tilde{\mathbf{y}}\right)\right) \tag{16}$$

To summarise, given independent scalar measurements $\mathbf{y}_{1:K}$ with additive ST $\left(\xi, \sigma^2, \lambda, \nu\right)$ noise, and a multivariate normal prior with mean $\mathbf{m}_0$ and covariance $\mathbf{P}_0$, the EM algorithm computes the MAP estimate $\hat{\mathbf{x}}$ as shown in Algorithm 1.

Algorithm 1 Computing position estimate by EM

Input: $\mathbf{y}_{1:K}$, $\mathbf{m}_0$, $\mathbf{P}_0$, n_{EM} (number of EM iterations) and n_{dGN} (number of descending GN iterations)
Initialise $\mathbf{x}^{(0)} \leftarrow \mathbf{m}_0$, $t_{1:K}^{(0)} \leftarrow -\xi/\delta_\lambda$ and $\tau_{1:K}^{(0)} \leftarrow 1$
for $j = 1$ to n_{EM} **do**
  Given $t_{1:K}^{(j-1)}$ and $\tau_{1:K}^{(j-1)}$, initialise $\hat{\mathbf{x}} \leftarrow \mathbf{x}^{(j-1)}$, and compute modified "data" $\tilde{\mathbf{y}} = \mathbf{y}_{1:K} - \xi - \delta_\lambda t_{1:K}^{(j-1)}$ and $\mathbf{R} = \sigma^2(1-\delta_\lambda^2)\mathrm{diag}\{1/\tau_1^{(j-1)}, .., 1/\tau_K^{(j-1)}\}$
  for $i = 1$ to n_{dGN} **do**
    Compute $\mathbf{H}$ using (15a)
    Compute $\mathbf{K} = \mathbf{P}_0\mathbf{H}^T\left(\mathbf{R} + \mathbf{H}\mathbf{P}_0\mathbf{H}^T\right)^{-1}$ and $d_{\mathrm{GN}} = \mathbf{m}_0 - \hat{\mathbf{x}} + \mathbf{K}\left(\tilde{\mathbf{y}} - \mathbf{h}(\hat{\mathbf{x}}) - \mathbf{H}\mathbf{m}_0 + \mathbf{H}\hat{\mathbf{x}}\right)$ (or $d_{\mathrm{GN}} = \left(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}\left(\mathbf{h}(\hat{\mathbf{x}}) - \tilde{\mathbf{y}}\right)$ if $\mathbf{P}_0^{-1} = \mathbf{0}$, i.e. if prior is flat)
    Compute $f(\hat{\mathbf{x}})$ using (21) with $\mathbf{x} \leftarrow \hat{\mathbf{x}}$
    Set $\alpha \leftarrow 1$ and compute $f(\hat{\mathbf{x}} + \alpha d_{\mathrm{GN}})$ using (21) with $\mathbf{x} \leftarrow \hat{\mathbf{x}} + \alpha d_{\mathrm{GN}}$
    while $f(\hat{\mathbf{x}} + \alpha d_{\mathrm{GN}}) \geq f(\hat{\mathbf{x}})$ **do**
      Set $\alpha \leftarrow \alpha/2$, and compute $f(\hat{\mathbf{x}} + \alpha d_{\mathrm{GN}})$ using (21) with $\mathbf{x} \leftarrow \hat{\mathbf{x}} + \alpha d_{\mathrm{GN}}$
    end while
    Set $\hat{\mathbf{x}} \leftarrow \hat{\mathbf{x}} + \alpha d_{\mathrm{GN}}$
  end for
  Set $\mathbf{x}^{(j)} \leftarrow \hat{\mathbf{x}}$
  Given $\mathbf{x}^{(j)}$, compute $t_{1:K}^{(j)}$ and $\tau_{1:K}^{(j)}$ using (12)
end for

B. Details of the EM algorithm

In this subsection the equations of the EM presented above are derived and details of the descending Gauss-Newton algorithm are given. Assuming a multivariate-normal prior distribution for $\mathbf{x}$ with mean $\mathbf{m}_0$ and covariance $\mathbf{P}_0$ and given the priors for the hyperparameters, the joint prior distribution is

$$\begin{aligned} p(\mathbf{x}, t_{1:K}, \tau_{1:K}) &\propto p_{\underline{\mathbf{x}}}(\mathbf{x}) p_{\underline{t}_{1:K}|\underline{\tau}_{1:K}}(t_{1:K}|\tau_{1:K}) p_{\underline{\tau}_{1:K}}(\tau_{1:K}) \\ &= \mathrm{e}^{-\frac{1}{2}(\mathbf{x}-\mathbf{m}_0)'\mathbf{P}_0^{-1}(\mathbf{x}-\mathbf{m}_0)} \prod_{k=1}^K [t_k \geq 0] \\ &\quad \mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{k=1}^K \tau_k t_k^2} \prod_{k=1}^K \tau_k^{\frac{\nu}{2}-1}\mathrm{e}^{-\frac{2}{\nu}\tau_k}. \end{aligned} \tag{17}$$

Thus, the joint posterior is

$$\begin{aligned} p(\mathbf{x}, t_{1:K}, \tau_{1:K}|\mathbf{y}_{1:K}) &\propto \mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{k=1}^K \frac{\tau_k}{1-\delta_\lambda^2}(\bar{y}_k - \delta_\lambda t_k)^2} \mathrm{e}^{-\frac{1}{2}(\mathbf{x}-\mathbf{m}_0)'\mathbf{P}_0^{-1}(\mathbf{x}-\mathbf{m}_0)} \\ &\quad \prod_{k=1}^K [t_k \geq 0] \mathrm{e}^{-\frac{1}{2\sigma^2}\sum_{k=1}^K \tau_k t_k^2} \prod_{k=1}^K \tau_k^{\frac{\nu}{2}-1}\mathrm{e}^{-\frac{2}{\nu}\tau_k}. \end{aligned} \tag{18}$$

Denoting $\tau_{-k} = \{\tau_1, .., \tau_{k-1}, \tau_{k+1}, .., \tau_K\}$ the conditional distribution of τ_k is

$$\begin{aligned} p(\tau_k|\mathbf{x}, \mathbf{y}_{1:K}, t_{1:K}, \tau_{-k}) &\propto \mathrm{e}^{-\frac{\tau_k}{2\sigma^2\left(1-\delta_\lambda^2\right)}(\bar{y}_k - \delta_\lambda t_k)^2} \tau_k^{\frac{\nu}{2}-1}\mathrm{e}^{-\frac{2}{\nu}\tau_k} \\ &= \tau_k^{\frac{\nu}{2}-1}\mathrm{e}^{-\tau_k\frac{(\bar{y}_k - \delta_\lambda t_k)^2 + 4\sigma^2\left(1-\delta_\lambda^2\right)}{2\sigma^2\left(1-\delta_\lambda^2\right)\nu}}, \end{aligned} \tag{19}$$

which is a Gamma distribution with shape parameter $\nu/2$ and scale parameter $\frac{2\sigma^2\left(1-\delta_\lambda^2\right)\nu}{(\bar{y}_k - \delta_\lambda t_k)^2 + 4\sigma^2\left(1-\delta_\lambda^2\right)}$. The mean value of this Gamma distribution, which is the product of shape and scale, is used in (12a) to update τ_k.

Denoting $t_{-k} = \{t_1, .., t_{k-1}, t_{k+1}, .., t_K\}$ the conditional

TABLE I

Number of operations from different classes for descending GN and EM algorithm, dependent on number of measurements K, number of EM iterations n_{EM} and number of descending GN iterations n_{dGN}.

Class	dGN	EM
addition	$\mathcal{O}(n_{\text{dGN}}K^3)$	$\mathcal{O}(n_{\text{EM}}n_{\text{dGN}}K^3)$
subtraction	$\mathcal{O}(n_{\text{dGN}}K)$	$\mathcal{O}(n_{\text{EM}}n_{\text{dGN}}K)$
multiplication	$\mathcal{O}(n_{\text{dGN}}K^3)$	$\mathcal{O}(n_{\text{EM}}n_{\text{dGN}}K^3)$
division	$\mathcal{O}(n_{\text{dGN}}K)$	$\mathcal{O}(n_{\text{EM}}n_{\text{dGN}}K)$
other	$\mathcal{O}(n_{\text{dGN}}K)$	$\mathcal{O}(n_{\text{EM}}n_{\text{dGN}}K)$

distribution of t_k is

$$\begin{aligned} p(t_k|\mathbf{x},\mathbf{y}_{1:K},t_{-k},&\tau_{1:K}) \\ &\propto \mathrm{e}^{-\frac{\tau_k}{2\sigma^2\left(1-\delta_\lambda^2\right)}\left(\bar{y}_k-\delta_\lambda t_k\right)^2}[t_k\geq 0]\mathrm{e}^{-\frac{\tau_k}{2\sigma^2}t_k^2} \\ &= [t_k\geq 0]\mathrm{e}^{-\frac{\tau_k}{2\sigma^2\left(1-\delta_\lambda^2\right)}\left(\left(\frac{\bar{y}_k}{\delta_\lambda}-t_k\right)^2+t_k^2\right)} \\ &\propto [t_k\geq 0]\mathrm{e}^{-\frac{1}{2\frac{\sigma^2\left(1-\delta_\lambda^2\right)}{2\tau_k}}\left(t_k-\frac{\bar{y}_k}{2\delta_\lambda}^2\right)}, \end{aligned} \quad (20)$$

which is a truncated normal distribution, i.e. values smaller zero are not allowed (due to $[t_k \geq 0]$), with center $\mu_k = \frac{1}{2\delta_\lambda}\left(y_k - h_k(\mathbf{x}) - \xi\right)$ and scale $\sigma_k = \sqrt{\frac{\sigma^2\left(1-\delta_\lambda^2\right)}{2\tau_k}}$. The mean value of this distribution is used in (12b) to update t_k.

The α used in (15b) and (16) ensures that the cost function

$$\begin{aligned} f(\mathbf{x}) \quad = \quad & \frac{1}{2}\Big((\mathbf{x}-\mathbf{m}_0)'\mathbf{P}_0^{-1}(\mathbf{x}-\mathbf{m}_0) \\ & + \left(\mathbf{h}_{1:K}(\mathbf{x})-\tilde{\mathbf{y}}\right)^T\mathbf{R}^{-1}\left(\mathbf{h}_{1:K}(\mathbf{x})-\tilde{\mathbf{y}}\right)\Big) \end{aligned} \quad (21)$$

does not increase, which is possible in the standard GN algorithm which uses the step $d_{\text{GN}} = \mathbf{m}_0 - \hat{\mathbf{x}} + \mathbf{K}\left(\tilde{\mathbf{y}} - \mathbf{h}(\hat{\mathbf{x}}) - \mathbf{H}\mathbf{m}_0 + \mathbf{H}\hat{\mathbf{x}}\right)$. After initialising $\alpha \leftarrow 1$ its value is repeatedly halved as long as $f(\hat{\mathbf{x}} + \alpha d_{\text{GN}}) \geq f(\hat{\mathbf{x}})$.

Table I shows a detailed complexity analysis of the EM method and compares it with the complexity of the descending GN, which will be used in Section IV as reference. The EM algorithm has for each operation class a n_{EM} times higher computational complexity. The number of operations in the E-step, for updating the hyperparameters, is negligible small in comparison to the number of operations in the M-step (the descending GN).

IV. Simulation experiment

In this section the positioning performance of using the skew-t likelihood model instead of a normal distribution likelihood model when the measurement errors are skew-t distributed is analysed. The MAP estimate for the normal model is computed using the descending GN algorithm. The source code and the full test suite are available at `https://PMullerTUT@bitbucket.org/PMullerTUT/trilaterationskewterrors.git`.

Here $\underline{\mathbf{x}}$ is a two-dimensional position and is assumed to have the prior distribution

$$\mathbf{x} \sim \text{MVN}\left(\mathbf{m}_0, \mathbf{P}_0\right) = \text{MVN}\left(\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right], 100\,\mathrm{I}_{2\times 2}\right), \quad (22)$$

TABLE II

Positioning error statistics for simulations with additive skew-t noises. Column *Time* gives the relative computation time using a specific MATLAB implementation, scaled so that computation time for the descending GN is 1.

Method	Time	Mean	Median	95 perc.
dGN	1	1.53	1.39	2.97
EM	5	1.15	1.05	2.59

where MVN(.,.) denotes a bivariate Gaussian distribution with given mean and covariance matrix. Four reference nodes are located at the corners of a 40-by-40 square centred at $\mathbf{m}_0$.

For the experiment 100 target positions are drawn from the prior distribution (22). Using (3) for computing the true distance between the target position and the reference node, $K = 12$ independent distance measurements (three to each reference node) are drawn from

$$y_k|\mathbf{x} \sim \text{ST}\left(\xi + h_k(\mathbf{x}), \sigma^2, \lambda, \nu\right) = \text{ST}\left(2 + h_k(\mathbf{x}), 3^2, 3, 3\right). \quad (23)$$

The EM algorithm uses the hierarchical version (11) of the measurement model with initial values $\tau_k = 1$ and $t_k = -\xi/\delta_\lambda$, which ensures that the descending GN finds the minimiser of the likelihood for Gaussian noise. For the EM algorithm 4 iterations are performed, and in each M-step 4 iterations of the descending GN algorithm are performed.

For comparison a descending GN that assumes measurement errors to be distributed as

$$y_k|\mathbf{x} \sim \text{N}\left(\xi + \sigma g_\nu \delta_\lambda + h_k(\mathbf{x}), \sigma^2\left(\frac{\nu}{\nu-2} - (g_\nu\delta_\lambda)^2\right)\right) \quad (24)$$

with 4 iterations is used. The parameters given in (24) ensure that the normal distribution has the same mean and variance as the skew-t distribution used by the EM algorithm.

In both the EM and the descending GN algorithm the number of repetitions to find a suitable α for (15b) and (16) is limited to 5 [4].

Table II presents the error statistics for the algorithms. *Mean* is the empirical mean, *Median* is the empirical median and *95% err* is the 95th percentile of all two-dimensional positioning errors, which are defined as the Euclidean distance (compare (3)) between the true position $\mathbf{x}$ and the position estimate $\hat{\mathbf{x}}$. *Time* gives the relative computation time using a specific MATLAB implementation, scaled so that computation time for the descending GN is 1. Fig. 1 shows the first 50 simulated positions and the corresponding estimates by EM and descending GN algorithm.

The EM algorithm clearly outperforms the descending GN algorithm in all three accuracy measures. That improvement in precision comes at the cost of an approximately five times higher computation time, which was expected since the EM uses four times more descending GN iterations and uses in addition the E-step for updating the hyperparameters. However, the computing cost is still reasonable, and could be brought down by tweaking the algorithm code and parameters.

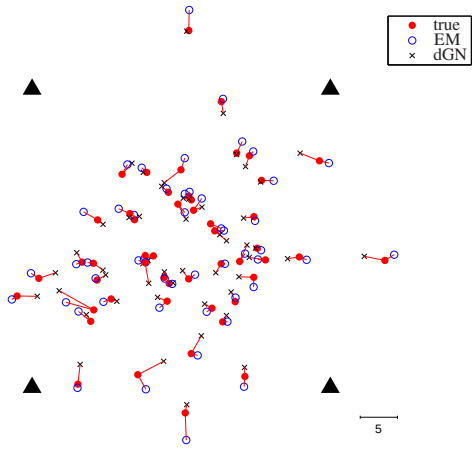


Fig. 2. First 50 simulated positions and their corresponding estimates computed by the EM and the descending GN algorithms.

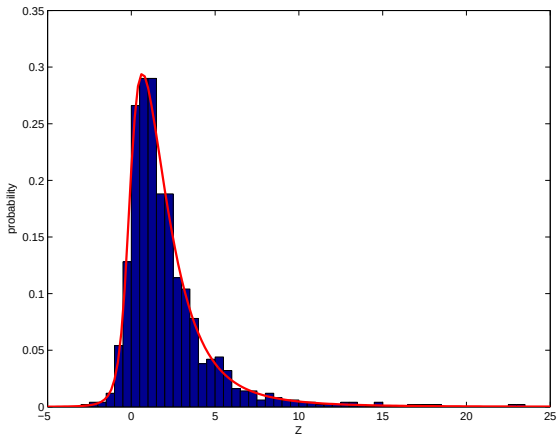


Fig. 3. Histogram of training data drawn from $v_j \sim \mathrm{ST}\left(0, 3^2, 3, 3\right)$ and pdf of a skew-t distribution that uses the median values given in Table III for $T_0 = 1\,000$ and $T = 10\,000$.

V. Fitting parameters of a skew-t distribution

A. Fitting parameters using a Gibbs sampler

In Section IV the error distribution's parameters are assumed to be known. For real-world applications it will be necessary to estimate the parameters by fitting them to a set of training data. In the trilateration setting (Section IV), training data are typically obtained in a measurement campaign in which ranges observed from known locations are collected, which allows determination of ranging errors. This section briefly describes the Gibbs sampling algorithm [6], which can be used to compute statistics of the posterior distributions for the parameters from n independent ranging errors $v_j|\xi, \sigma^2, \lambda, \nu \sim \mathrm{ST}\left(\xi, \sigma^2, \lambda, \nu\right)$.

The Gibbs sampling method could also be used to compute position, but can be expected to be much slower than EM.

The idea of the Gibbs sampler (GS) is to sample from the conditional posterior distributions for each parameter separately when sampling from the (multivariate) posterior is not feasible. For the parameter estimation problem at hand generating samples from the joint posterior $p\left(\xi, \sigma^2, \lambda, \nu|v_{1:n}\right)$ is unfeasible, but sampling from the conditional posterior distributions $p\left(\xi|\sigma^2, \lambda, \nu, v_{1:n}\right)$, $p\left(\sigma^2|\xi, \lambda, \nu, v_{1:n}\right)$, $p\left(\lambda|\xi, \sigma^2, \nu, v_{1:n}\right)$ and $p\left(\nu|\xi, \sigma^2, \lambda, v_{1:n}\right)$ is possible.

The algorithm works as follows. First, initial values $\xi_{(0)}, \sigma^2_{(0)}, \lambda_{(0)}, \nu_{(0)}$ are assigned to the unknown parameters. Then the parameters are ordered and samples from the conditional distribution of each parameter given the error data $v_{1:n}$ and the current estimates of the remaining parameters are drawn. This updating process is repeated $T_0 + T$ times. The posterior means of the estimates are estimated by the empirical sample means of the last T samples; the first T_0 "burn-in" samples are discarded. For example,

$$\mathrm{E}\left(\hat{\xi}|v_{1:n}\right) \approx \frac{1}{T}\sum_{t=1}^{T} \xi_{(t+T_0)}. \tag{25}$$

Other statistics, such as the posterior median, can be estimated in a similar way.

Since in the problem under consideration four parameters have to be estimated a large set of training data will be required to obtain estimates that are close to the true values of the parameters. If priors are vague, a large training set is needed to obtain a posterior with small dispersion. A smaller training set can be used if the priors are more informative.

B. An example for fitting parameters by the GS

In this subsection a small demo for fitting parameters of a skew-t distribution using the GS algorithm is presented. The training data consists of $n = 1\,000$ observations that are drawn from

$$v_j|\xi = 0, \sigma^2 = 9, \lambda = 3, \nu = 3 \sim \mathrm{ST}\left(\xi, \sigma^2, \lambda, \nu\right). \tag{26}$$

The source code and the full test suite are available at `https://PMullerTUT@bitbucket.org/PMullerTUT/trilaterationskewterrors.git`.

For estimating the parameters a hierarchical model similar to (11) and JAGS [7], which is a program that allows analysing Bayesian hierarchical models by Gibbs samplers, are used. Algorithm 2 shows the pseudo-code for the model used in JAGS.

The GS is run twice with different $\xi_{(0)}, \sigma^2_{(0)}, \lambda_{(0)}, \nu_{(0)}$. For each parameter a diffuse prior is used; only for ν a slightly informative prior, namely a uniform distribution over the interval $(2, 100)$ is used because $\nu > 2$ is required in (6).

Table III contains the summary statistics of the posterior distribution for different numbers of "burn-in" and retained

Algorithm 2 Model for JAGS used for fitting parameters

Input: number of measurements n
Draw parameters from priors: $\xi \sim \mathrm{N}\left(0, 100^2\right)$, $\sigma = 1/p$ with $p \sim \Gamma\left(10\,000, 10\,000\right)$, $\lambda \sim \mathrm{N}\left(0, 100^2\right)$ and $\nu \sim \mathrm{Uniform}\left(2, 100\right)$
Compute $\delta_\lambda = \frac{\lambda}{\sqrt{1+\lambda^2}}$
for $j = 1$ to n **do**
  draw $\tau_j \sim \Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$
  draw $t_j \sim \mathrm{HN}\left(0, \frac{\sigma^2}{\tau_j}\right)$ (see (7) for its pdf)
  draw $v_j \sim \mathrm{N}\left(\xi + \delta_\lambda t_j, \frac{1-\delta_\lambda^2}{\tau_j}\sigma^2\right)$
end for
draw predicted latent variable $\tau^{(\mathrm{pr})} \sim \Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$
draw predicted latent variable $t^{(\mathrm{pr})} \sim \mathrm{HN}\left(0, \frac{\sigma^2}{\tau^{(\mathrm{pr})}}\right)$
draw predicted observation $v^{(\mathrm{pr})} \sim \mathrm{N}\left(\xi + \delta_\lambda t^{(\mathrm{pr})}, \frac{1-\delta_\lambda^2}{\tau^{(\mathrm{pr})}}\sigma^2\right)$

TABLE III
Statistics of posterior distributions estimated from 1 000 observations, drawn from $v_j|\xi = 0, \sigma^2 = 9, \lambda = 3, \nu = 3 \sim \mathrm{ST}\left(\xi, \sigma^2, \lambda, \nu\right)$, by two Gibbs samplers with each T_0 "burn-in" and T retained samples.

Parameter	T_0	T	5%-ile	Median	95%-ile
ξ	200	500	-0.230	-0.114	-0.001
σ^2	200	500	3.474	4.129	5.065
λ	200	500	2.867	3.681	4.617
ν	200	500	2.390	2.818	3.410
ξ	1000	10000	-0.219	-0.115	-0.008
σ^2	1000	10000	3.455	4.159	5.001
λ	1000	10000	2.987	3.646	4.459
ν	1000	10000	2.390	2.813	3.389

samples. For each parameter the median, and 5%-ile and 95%-ile are given.

Already with $T_0 = 200$ "burn-in" and $T = 500$ retained samples the parameter estimates are quite good; only σ^2 is underestimated slightly. Using $T_0 = 1\,000$ and $T = 10\,000$ does not improve the quality of the estimates significantly. To further improve the estimates more training data (i.e. observations) would be required or could be fixed to reasonable values and the GS could be repeatedly run with various fixed ν for estimating ξ, σ^2 and λ. However, from Fig. 3, which shows the histogram of the training data and the pdf of a skew-t distribution that uses the median values given in Table III for $T_0 = 1\,000$ and $T = 10\,000$, these approaches seem to be unnecessary.

The computation times on a laptop where 15 seconds for $T_0 = 200$ and $T = 500$ and less than 4 minutes for $T_0 = 1\,000$ and $T = 10\,000$, which is tolerable for an algorithm that is intended for offline use.

As mentioned in the introduction, the non-line-of-sight data from UWB network in [2] is clearly skewed. Fig. 4 shows the density histogram of the data and a skew-t distribution fitted by JAGS using $T_0 = 1\,000$ and $T = 10\,000$ and two GS.

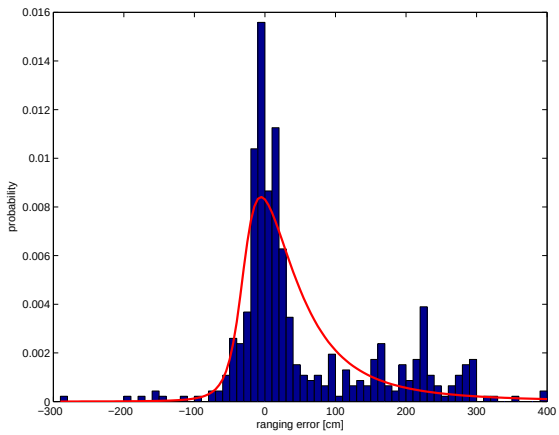


Fig. 4. Density histogram of non-line-of-sight errors from UWB data in [2] and fitted skew-t $\mathrm{ST}\left(-30.73, 4450.05, 3.22, 2.05\right)$.

VI. Concluding Remarks

This paper explained how the trilateration problem can be solved using an Expectation-Maximisation algorithm when the measurements contain additive skew-t distributed errors. For such measurement data it was shown by simulations that the presented EM algorithm improves the positioning precision significantly compared with a descending Gauss-newton algorithm, which models the measurement noise as being additive normal. This improvement in precision comes at the cost of higher computational demand.

In addition, it is shown how the parameters of a skew-t can be fitted to training data by using a Gibbs sampler.

Future work includes the testing of the algorithm with real-world data that has skew measurement noise. Moreover, the algorithm can be modified for other measurement functions, e.g. pseudo-ranges, angle-of-arrival or time-of-arrival type measurements, and the algorithm could be implemented in a filter to further improve its positioning precision.

References

[1] N. Sirola, "Closed-form algorithms in mobile positioning: myths and mis- conceptions," in *Proceedings of the 7th Workshop on Positioning, Navigation and Communications*, 2010, pp. 38–44.

[2] P. Müller, H. Wymeersch, and R. Piché, "UWB positioning with generalized Gaussian mixture filters," *IEEE Transactions on Mobile Computing*, vol. 13, no. 10, pp. 2406–2414, October 2014.

[3] A. Azzalini, *The skew-normal and related families*. Cambridge University Press, 2014.

[4] R. Piché, "Estimation of model parameters," in *Mathematical Modeling with Multidisciplinary Applications*, X.-S. Yang, Ed. John Wiley and Sons, 2013, pp. 169–190.

[5] T. I. Lin, J. C. Lee, and W. J. Hsieh, "Robust mixture modeling using the skew t distribution," *Statistics and Computing*, vol. 17, no. 2, pp. 81–92, June 2007.

[6] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721–741, 1984.

[7] (2013, September). [Online]. Available: http://mcmc-jags.sourceforge.net

UNPUBLISHED MANUSCRIPT 1

Philipp Müller, José A. del Peral-Rosado, Robert Piché, and Gonzalo Seco-Granados: Statistical Trilateration with Skew-t Errors in LTE networks.

Statistical Trilateration with Skew-t Distributed Errors in LTE Networks

Philipp Müller, José A. del Peral-Rosado, Robert Piché, and Gonzalo Seco-Granados

***Abstract*—Localization accuracy of trilateration methods in Long Term Evolution (LTE) cellular networks may be highly degraded due to multipath and non-line of sight (NLoS) conditions in urban and indoor environments. Multipath mitigation techniques usually involve a high computational burden and wideband signals to be effective, which limit their adoption in certain mobile applications. As an alternative to these conventional techniques, this paper analyzes an Expectation Maximization (EM) localization algorithm that considers the skewness introduced by multipath in the LTE ranging measurements. The EM algorithm is extensively studied with realistic emulated LTE signals of 1.4 MHz bandwidth. The EM method is compared to a standard Nonlinear Least Squares (NLS) algorithm under ideal simulated conditions and using realistic measurements from a laboratory testbed. The EM method outperforms the NLS method when the ranging errors in the training and test stages have similar distributions.**

***Index Terms*—Statistical trilateration, skew-t distribution, Localization, LTE, Expectation Maximization algorithm**

I. Introduction

MOBILE phone localization has been typically supported by Global Navigation Satellite Systems (GNSS) or network-based methods with coarse accuracy, such as cell identification (CID). But the worldwide adoption of the Long Term Evolution (LTE) in mobile networks is starting to change this paradigm. Complementary methods based on the trilateration of LTE pilot signals are specified in the standard [1] by the Third Generation Partnership Project (3GPP), such as the observed time difference of arrival (OTDoA) and the uplink time difference of arrival (UTDoA). These trilateration methods estimate the receiver position based on measurements of distances (ranges) between the receiver and base stations (BSs) with known location. The use of pilot signals with up to 20 MHz of bandwidth specified in LTE [2], e.g. cell-specific reference signal (CRS), positioning reference signal (PRS) or sounding reference signal (SRS), is of special interest for accurate positioning in urban areas, where GNSS signal availability is highly reduced.

Thanks to the tight network synchronization, the main source of positioning error using trilateration techniques in LTE is due to multipath and non-line of sight (NLoS) conditions. These effects can be mitigated with techniques to improve the ranging measurements (e.g. [3]–[5]). The most typical techniques are based on the correlation of the received signal with a pilot signal and the detection of the first peak above a threshold, in order to compute the time-delay estimation (TDE) or range estimation. These techniques, called threshold-based estimators [3], have a good performance for moderate to high signal bandwidth, such as the common LTE system bandwidths of 5 and 10 MHz [6]. However, the multipath mitigation capabilities of these techniques is significantly decreased for low signal bandwidths, such as 1.4 MHz, where multipath reflections are not individually distinguishable. For instance, the demand of high data rates in fourth-generation (4G) networks is expected to limit the allocated bandwidth for positioning services. This context leads to the challenging goal of finding positioning techniques able to achieve accurate LTE localization in multipath environments for the lowest system bandwidth, i.e., by using only pilot signals over 1.4 MHz.

A complementary approach to multipath countermeasures in the ranging estimation, as in [7], [8], is to make use of a more elaborate statistical signal model in the localization algorithm. Multipath channels introduce skewness to the error distribution of time-of-arrival (ToA) or range measurements, as it has been observed in [9] and [10] for ultra-wideband (UWB) networks. Thus, taking into account this skewed or asymmetric distribution in the localization algorithm may give better performance than a standard Nonlinear Least Squares (NLS) algorithm (which is implicitly based on a Gaussian error model).

Many different non-Gaussian models of range error have been proposed, including, among others, the exponential (e.g. in [5] for NLoS errors), log-normal, Weibull, generalized extreme value [11], normal-Cauchy mixture [10], and normal-exponential mixture [12] distributions. In [13] we considered the skew-t distribution as a range error model and presented an Expectation Maximization (EM) algorithm for trilateration. The paper showed how the parameters of a skew-t distribution can be fitted to training data using a Gibbs sampler (GS). Furthermore, in that paper simulations in idealized trilateration models showed significant positioning accuracy improvements over a standard NLS algorithm.

The motivation of this paper is to make a thorough study

P. Müller and R. Piché are with the Department of Automation Science and Engineering (ASE), Tampere University of Technology, P.O. Box 692, FI-33101 Tampere, Finland.
E-mail: philipp.muller@tut.fi, robert.piche@tut.fi

J. del Peral-Rosado and G Seco-Granados are with Department of Telecommunications and Systems Engineering, Universitat Autònoma de Barcelona, Q Building, Cerdanyola del Vallès 08193, Spain.
E-mail: JoseAntonio.DelPeral@uab.cat, Gonzalo.Seco@uab.cat

P. Müller acknowledges the financial support of the Doctoral Programme of TUT's President, and the financial support by the Finnish Doctoral Programme in Computational Sciences (FICS) and TUT-Foundation (TTY-säätiö) for his research visit to Universitat Autònoma de Barcelona, Spain. Part of this work was supported by the European Space Agency (ESA) under the NPI programme No. 4000110780/14/NL/AK, and by the Spanish Ministry of Economy and Competitiveness project TEC2014-53656-R.

of the applicability of skew-t range error modelling approach of [13] in a realistic trilateration setting, and in particular, to evaluate it as a multipath mitigation technique to LTE positioning effective for the lowest signal bandwidth, where multipath components are unresolvable. In this paper we fit skew-t distributions to training data from 3GPP standard channel models using the Gibbs sampler presented in [13], and analyze the EM positioning accuracy in an LTE network using ranging measurements obtained from a laboratory testbed. The assessment of this statistical trilateration technique is performed under harsh conditions by using emulated data and the lowest system bandwidth of LTE, i.e., 1.4 MHz. Thus, the practical performance of the skew-t trilateration algorithm is validated in realistic conditions.

This paper is organized as follows. In Section II we provide a review of the essential aspects of the trilateration problem, the skew-t distribution and the fitting to experimental data using a GS, as well as a description of LTE channel models and ranging methods. Although the contents of Sections II.A and II.B can also be mostly found in [13], we consider it pertinent to include them here to put the rest of the paper in context and make it understandable. In Section III, we present briefly the EM algorithm from [13] used for solving the trilateration problem assuming skew-t distributed errors and the dilution of precision. We then apply in Section IV the algorithms for positioning in an LTE network. After describing the testbed, we fit parameters of skew-t and normal distributions to ranging errors from LTE standard channel models. Subsequently, we test the EM algorithm positioning performance for both channels and compare it with the performance of a standard NLS solver. Finally, we give some concluding remarks and an outlook in Section V.

Notation: $\mathbf{x}$ and $x_{1:d}$ denote column vectors, $\mathbf{H}$ denotes a matrix, and underscores are used to denote random variables and random vectors in contexts where the distinction from deterministic variables is useful.

II. Model for statistical trilateration in LTE

A. Statistical trilateration

In this paper we use the following statistical formulation of the trilateration problem [14]. Let the d-dimensional random vector $\underline{x}_{1:d} = \underline{\mathbf{x}}$ represent the unknown receiver location. The K scalar measurements are modeled as

$$\underline{y}_k | \left(\underline{\mathbf{x}} = \mathbf{x}\right) = h_k(\mathbf{x}) + \underline{v}_k \tag{1}$$

for $k \in 1, .., K$, where function $h_k : \mathbb{R}^d \to \mathbb{R}$ models the measurement geometry, and the additive errors $\underline{v}_1, .., \underline{v}_K$ are mutually independent random variables. Furthermore, $\underline{v}_{1:K}$ and $\underline{\mathbf{x}}$ are independent from each other.

The prior probability density function (pdf) of $\mathbf{x}$ is denoted as $p_{\underline{\mathbf{x}}}$, and the pdf of v_k as $p_{\underline{v}_k}$. The posterior distribution of $\mathbf{x}$ given the K-dimensional measurement vector $y_{1:K} = \mathbf{y}$ has the pdf

$$p_{\underline{\mathbf{x}}|\underline{\mathbf{y}}}(\mathbf{x}|\mathbf{y}) \propto p_{\underline{\mathbf{x}}}(\mathbf{x}) \prod_{k=1}^{K} p_{\underline{v}_k}\left(y_k - h_k(\mathbf{x})\right), \tag{2}$$

where $\propto$ means "proportional to". A value of $\mathbf{x}$ that maximizes (2) is called a maximum *a posteriori* (MAP) estimate. If the prior distribution is "flat", i.e. if $p_{\underline{\mathbf{x}}}(\mathbf{x}) \propto 1$, this MAP estimate coincides with the maximum likelihood (ML) estimate.

In trilateration, the measurement function is the Euclidean distance between the receiver and a base station at a known location $\mathbf{c}_k$:

$$h_k(\mathbf{x}) = ||\mathbf{x} - \mathbf{c}_k||. \tag{3}$$

The Jacobian of $\mathbf{h} = h_{1:K}$ is the $K \times d$ matrix $\mathbf{H}$ whose kth row is the transpose of a unit vector pointing from $\mathbf{c}_k$ to $\mathbf{x}$, that is,

$$\mathbf{H}_{k,1:d} = \frac{\partial h_k(\mathbf{x})}{\partial \mathbf{x}} = \frac{\left(\mathbf{x} - \mathbf{c}_k\right)^T}{||\mathbf{x} - \mathbf{c}_k||}. \tag{4}$$

B. The skew-t distribution

This subsection presents the skew-t distribution and some of its properties that are needed later in this paper. For a more extensive discussion we refer the reader, for example, to [15, pp. 101 ff.] and [16].

A random variable $\underline{z}$ is said to have a skew-t distribution with location ξ, scale σ^2, skewness λ and ν degrees-of-freedom (dof) if its pdf has the form

$$p_{\underline{z}}(z) = \frac{2}{\sigma} t_\nu\left(\frac{z-\xi}{\sigma}\right) T_{\nu+1}\left(\lambda \frac{z-\xi}{\sigma}\sqrt{\frac{\nu+1}{\nu + \frac{(z-\xi)^2}{\sigma^2}}}\right), \tag{5}$$

where t_ν and $T_{\nu+1}$ denote the pdf and the cumulative distribution function (cdf) of the standardized t-distribution. The skew-t distribution is denoted $\underline{z} \sim \mathrm{ST}\left(\xi, \sigma^2, \lambda, \nu\right)$.

For $\nu > 2$ the mean and variance of $\underline{z} \sim \mathrm{ST}\left(\xi, \sigma^2, \lambda, \nu\right)$ are

$$\mathrm{E}(\underline{z}) = \xi + \sigma g_\nu \delta_\lambda, \quad \mathrm{var}(\underline{z}) = \sigma^2\left(\tfrac{\nu}{\nu-2} - (g_\nu \delta_\lambda)^2\right) \tag{6}$$

where $g_\nu = \frac{\sqrt{\nu}\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}$ and $\delta_\lambda = \frac{\lambda}{\sqrt{1+\lambda^2}} \in (-1, 1)$.

The skew-t distribution has the following hierarchical model. Let $\underline{\tau} \sim \Gamma\left(\nu/2, \nu/2\right)$ and $\underline{w} \sim \mathrm{N}\left(0, 1\right)$, which denote the Gamma distribution with both shape and scale $\nu/2$ and the standard normal distribution. Then $\underline{t}|(\underline{\tau} = \tau) = |\sigma \underline{w}/\sqrt{\tau}|$ is a half-normal (HN) random variable with pdf

$$p_{\underline{t}}(t) = 2\frac{\sqrt{\tau}}{\sigma}\phi\left(t\frac{\sqrt{\tau}}{\sigma}\right)[t > 0], \tag{7}$$

where $[\,\cdot\,]$ is the Iverson bracket and ϕ denotes the pdf of the standard normal distribution. Samples from the distribution $\mathrm{ST}\left(\xi, \sigma^2, \lambda, \nu\right)$ can be drawn from the conditional distribution

$$\underline{z}|(\underline{t} = t, \underline{\tau} = \tau) \sim \mathrm{N}\left(\xi + \frac{\lambda t}{\sqrt{1+\lambda^2}}, \frac{1-\delta_\lambda^2}{\tau}\sigma^2\right). \tag{8}$$

In the hierarchical representation (7 – 8), the conditional random variable $\underline{t}|(\underline{z} = z, \underline{\tau} = \tau)$ has the distribution $\mathrm{N}\left(\delta_\lambda(z-\xi), \frac{1-\delta_\lambda^2}{\tau}\sigma^2\right)$ and the conditional random variable $\underline{\tau}|(\underline{z} = z)$ has the pdf

$$\underline{\tau}|(\underline{z} = z) \propto \tau^{(\nu-1)/2} \exp\left(-\frac{\tau}{2}(\eta^2 + \nu)\right) \Phi(\lambda\eta\sqrt{\tau}), \tag{9}$$

where $\eta = \frac{z-\xi}{\sigma}$.

C. Fitting skew-t parameters using a Gibbs sampler

For fitting the parameters of a skew-t distribution to training data we apply the Gibbs sampling algorithm [17]. This algorithm computes the statistics of the posterior distributions for the parameters from n independent ranging errors $e_j|\xi,\sigma^2,\lambda,\nu \sim \text{ST}\left(\xi,\sigma^2,\lambda,\nu\right)$.

The idea of the GS is to sample from the conditional posterior distributions for each parameter separately when sampling from the (multivariate) posterior is not feasible, which is the case in our parameter estimation problem.

The algorithm works as follows. First, initial values $\xi_{(0)},\sigma^2_{(0)},\lambda_{(0)},\nu_{(0)}$ are assigned to the unknown parameters. Then the parameters are ordered and samples from the conditional distribution of each parameter given the error data $e_{1:n}$ and the current estimates of the remaining parameters are drawn. This updating process is repeated T_0+T times. The posterior means of the estimates are estimated by the empirical sample means of the last T samples; the first T_0 "burn-in" samples are discarded. Other statistics, such as the covariance and quantiles, can be computed from the samples in a similar way.

For more details on how to use GS for skew-t parameter fitting we refer the reader to [13].

D. LTE ranging and channels

Current and next-generation mobile cellular networks are based on LTE systems, which are standardized by the 3GPP consortium. These systems are able to provide communications and positioning services within the same signal transmission. This is possible thanks to the use of multicarrier signals, which allow a flexible allocation of resources. The downlink transmission between BS and receiver or user equipment (UE) is based on the orthogonal frequency division multiplexing (OFDM) modulation. The LTE downlink also specifies the OTDoA trilateration method [1]. This positioning method uses time-delay estimates or ranging measurements performed with pilot or reference signals, such as CRS or PRS. The most common ranging estimator is based on the maximum likelihood (ML) approach as [7]

$$\hat{\tau} = \underset{\tau}{\text{argmax}} \left\{ \sum_{k\in\mathcal{K}} \left| \sum_{n\in\mathcal{N}} X\left(k,n\right)\cdot b^*\left(k,n\right)\cdot e^{-j\frac{2\pi n\tau}{N}} \right|^2 \right\}, \tag{10}$$

where τ is the time-delay, $X\left(k,n\right)$ is the frequency received signal at the k-th symbol and n-th subcarrier, $b^*\left(k,n\right)$ is the conjugate of the pilot code, $\mathcal{K}$ is the set of reference symbols, $\mathcal{N}$ is the set of reference subcarriers, and N is the total number of subcarriers.

LTE specifies OTDoA in [1] as a network-based trilateration method, where the receiver position is calculated at a location server. First, the receiver obtains the ranging measurements using assistance data provided by the network. Then, the location server requests the clock offsets and locations of each BS. Finally, the receiver position is computed and sent back to the receiver.

The 3GPP consortium defines the minimum performance requirements of the LTE standard. For this purpose, several propagation channel models have been specified for simulation and testing of the communication and positioning capabilities of this cellular system. The main channel models adopted in LTE are the Extended Pedestrian A (EPA), Extended Vehicular A (EVA) and Extended Typical Urban (ETU), which are intended to model multipath environments with low, medium and large delay spread, respectively [18]. These are tapped-delay line (TDL) models with time-varying channel coefficients with a Rayleigh fading distribution. Three-dimensional (3D) channel models are also specified in [19] for indoor environments. However, here we only consider the EPA and ETU channel models in order to evaluate the statistical trilateration in mild and harsh multipath conditions, respectively. Doppler shifts of 5 and 70 Hz are also considered in this paper to represent low and high mobility conditions, resulting in the EPA5 and ETU70 models, respectively, whose main parameters are shown in Table I, as specified in Annex B of [18]. In addition, only the lowest system bandwidth of LTE, i.e. 1.4 MHz, is used. At this bandwidth, the propagation rays jointly contribute to the multipath error on the ranging measurements, and they limit the efficiency of multipath mitigation techniques. Thus, positioning algorithms, such as statistical trilateration, that are able to mitigate these effects are of special interest.

TABLE I
Main parameters of the standard LTE channel models.

Tap	**EPA channel**		**ETU channel**	
no.	τ (ns)	SMR (dB)	τ (ns)	SMR (dB)
1	0	0.0	0	-1.0
2	30	-1.0	50	-1.0
3	70	-2.0	120	-1.0
4	90	-3.0	200	0.0
5	110	-8.0	230	0.0
6	190	-17.2	500	0.0
7	410	-20.8	1600	-3.0
8			2300	-5.0
9			5000	-7.0

III. Estimation of the receiver position

A. Positioning using Expectation Maximization

In this subsection, we present the basics of the EM algorithm for computing the MAP estimate $\hat{\mathbf{x}}$, i.e. the mode of (2), for additive skew-t measurement errors. The algorithm's maximization step (M-step) is the standard NLS problem. For details we refer the reader to [13].

Algorithm 1 shows the pseudo code of the method, which starts with the M-step. Given K independent scalar measurements $\mathbf{y}$ and assuming a multivariate-normal prior distribution for $\mathbf{x}$ with mean $\mathbf{m}$ and covariance $\mathbf{P}$, the M-step is the computation of the mode of the conditional posterior

$p(\mathbf{x}^{(j)}|\mathbf{y}, t_{1:K}^{(j-1)}, \tau_{1:K}^{(j-1)})$, that is,

$$\hat{\mathbf{x}} \leftarrow \underset{\mathbf{x}}{\operatorname{argmin}} \left((\mathbf{x}-\mathbf{m})' \mathbf{P}^{-1} (\mathbf{x}-\mathbf{m}) + \frac{\sum_{k=1}^{K} \tau_k^{(j-1)} \left(y_k - h_k(\mathbf{x}) - \xi - \delta_\lambda t_k^{(j-1)} \right)^2}{\sigma^2 (1-\delta_\lambda^2)} \right). \quad (11)$$

The minimization in (11) can be computed by any standard NLS algorithm; we use the descending Gauss-Newton (dGN) algorithm (see e.g. [14]). The dGN gets its name from the scaling factor α seen in Algorithm 1, which ensures a decreasing cost function

$$f(\mathbf{x}) = \frac{1}{2} \Big((\mathbf{x}-\mathbf{m})' \mathbf{P}^{-1} (\mathbf{x}-\mathbf{m}) + (\mathbf{h}(\mathbf{x}) - \tilde{\mathbf{y}})^T \mathbf{R}^{-1} (\mathbf{h}(\mathbf{x}) - \tilde{\mathbf{y}}) \Big), \quad (12)$$

where $\tilde{\mathbf{y}} = \mathbf{y} - \xi - \delta_\lambda t_{1:K}^{(j-1)}$ and $\mathbf{R} = \sigma^2(1-\delta_\lambda^2)\text{diag}\{1/\tau_1^{(j-1)}, .., 1/\tau_K^{(j-1)}\}$; and α is found by line search.

Using (8), a hierarchical version of the measurement model $\underline{y}_k|(\underline{\mathbf{x}} = \mathbf{x}) \sim \text{ST}\left(\xi + h_k(\mathbf{x}), \sigma^2, \lambda, \nu\right)$ is

$$\underline{y}_k|(\underline{\mathbf{x}} = \mathbf{x}, \underline{t}_k = t_k, \underline{\tau}_k = \tau_k) \sim \text{N}\left(\xi + h_k(\mathbf{x}) + \delta_\lambda t_k, \frac{1-\delta_\lambda^2}{\tau_k}\sigma^2\right), \quad (13)$$

where the hyperparameters are $\underline{t}_k|(\underline{\tau}_k = \tau_k) = |\sigma \underline{w}_k / \sqrt{\tau_k}|$ with $\underline{w}_k \sim \text{N}(0,1)$ and $\underline{\tau}_k \sim \Gamma(\nu/2, \nu/2)$. In the EM algorithm's expectation step (E-step) the hyperparameters are updated by setting them to the mean values of their conditional distributions. For $t_k^{(j)}$ the conditional distribution is a truncated normal distribution that rejects negative values, with center $\mu_k = \frac{1}{2\delta_\lambda}(y_k - h_k(\mathbf{x}) - \xi)$ and scale $\sigma_k = \sqrt{\frac{\sigma^2(1-\delta_\lambda^2)}{2\tau_k^{(j)}}}$; and for $\tau_k^{(j)}$ the conditional distribution is a Gamma distribution with shape parameter $\nu/2$ and scale parameter

$$\frac{2\sigma^2(1-\delta_\lambda^2)\nu}{\left(y_k - h_k(\mathbf{x}^{(j)}) - \xi - \delta_\lambda t_k^{(j-1)}\right)^2 + 4\sigma^2(1-\delta_\lambda^2)}.$$

In Algorithm 1, the update formula for t_k is written using the scaled complementary error function erfcx, a special function defined as $\exp(x^2)\text{erfc}(x)$ that is available in mathematical software libraries. The update formula presented in [13] is mathematically equivalent but gives significant numerical floating point computation errors for large negative values of μ_k/σ_k.

Since the trilateration problem in real-world applications usually has to be solved in real time in mobile devices, the EM algorithm rate of convergence is important. The EM algorithm is known to converge very slowly for some problems. However, Xu and Jordan [20] claim that for such problems also other gradient-based methods generally show slow convergence. They also state that the EM converges monotonically without requiring the user to set a learning rate, a property that not all alternative approaches have. Dempster et al. [21] show that the algorithm converges with a linear rate and that this rate depends on the amount of information inherent in the measurements. Our tests for [13] showed that for the trilateration problem the EM algorithm converges quickly; four iterations were enough to converge to the EM's position estimate.

Algorithm 1 Computing position estimate by EM

Input: $\mathbf{y}$, $\mathbf{m}$, $\mathbf{P}$, n_{EM} (number of EM iterations) and n_{dGN} (number of descending-GN iterations)
Initialize $\mathbf{x}^{(0)} \leftarrow \mathbf{m}$, $t_{1:K}^{(0)} \leftarrow -\xi/\delta_\lambda$ and $\tau_{1:K}^{(0)} \leftarrow 1$
for $j = 1$ to n_{EM} **do**
  Given $t_{1:K}^{(j-1)}$ and $\tau_{1:K}^{(j-1)}$, initialize $\hat{\mathbf{x}} \leftarrow \mathbf{x}^{(j-1)}$, and compute modified "data" $\tilde{\mathbf{y}} = \mathbf{y} - \xi - \delta_\lambda t_{1:K}^{(j-1)}$ and $\mathbf{R} = \sigma^2(1-\delta_\lambda^2)\text{diag}\{1/\tau_1^{(j-1)}, .., 1/\tau_K^{(j-1)}\}$
  for $i = 1$ to n_{dGN} **do**
    Compute $\mathbf{H} \leftarrow \frac{\partial \mathbf{h}_{1:K}}{\partial \mathbf{x}}(\hat{\mathbf{x}})$ using (4)
    Compute $\mathbf{K} = \mathbf{P}\mathbf{H}^T\left(\mathbf{R} + \mathbf{H}\mathbf{P}\mathbf{H}^T\right)^{-1}$ and $d_{\text{GN}} = \mathbf{m}_0 - \hat{\mathbf{x}} + \mathbf{K}(\tilde{\mathbf{y}} - \mathbf{h}(\hat{\mathbf{x}}) - \mathbf{H}\mathbf{m} + \mathbf{H}\hat{\mathbf{x}})$ (or $d_{\text{GN}} = \left(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{h}(\hat{\mathbf{x}}) - \tilde{\mathbf{y}})$ if $\mathbf{P}^{-1} \to \mathbf{0}$, i.e. if prior is flat
    Compute $f(\hat{\mathbf{x}})$ using (12) with $\mathbf{x} \leftarrow \hat{\mathbf{x}}$
    Set $\alpha \leftarrow 1$ and compute $f(\hat{\mathbf{x}} + \alpha d_{\text{GN}})$ using (12) with $\mathbf{x} \leftarrow \hat{\mathbf{x}} + \alpha d_{\text{GN}}$
    while $f(\hat{\mathbf{x}} + \alpha d_{\text{GN}}) \geq f(\hat{\mathbf{x}})$ **do**
      Set $\alpha \leftarrow \alpha/2$, and compute $f(\hat{\mathbf{x}} + \alpha d_{\text{GN}})$ using (12) with $\mathbf{x} \leftarrow \hat{\mathbf{x}} + \alpha d_{\text{GN}}$
    end while
    Set $\hat{\mathbf{x}} \leftarrow \hat{\mathbf{x}} + \alpha d_{\text{GN}}$
  end for
  Set $\mathbf{x}^{(j)} \leftarrow \hat{\mathbf{x}}$
  Given $\mathbf{x}^{(j)}$, compute $t_{1:K}^{(j)}$ and $\tau_{1:K}^{(j)}$ as follow
  $\tau_k^{(j)} \leftarrow \frac{\sigma^2(1-\delta_\lambda^2)\nu^2}{\left(y_k - h_k(\mathbf{x}^{(j)}) - \xi - \delta_\lambda t_k^{(j-1)}\right)^2 + 4\sigma^2(1-\delta_\lambda^2)}$
  $t_k^{(j)} \leftarrow \mu_k + \sigma_k\sqrt{\frac{2}{\pi}}/\text{erfcx}\left(\frac{-\mu_k/\sigma_k}{\sqrt{2}}\right)$
  where scaled complementary error function erfcx $= \exp(x^2)\text{erfc}(x)$, $\mu_k = \frac{1}{2\delta_\lambda}\left(y_k - h_k(\mathbf{x}^{(j)}) - \xi\right)$ and $\sigma_k = \sqrt{\frac{\sigma^2(1-\delta_\lambda^2)}{2\tau_k^{(j)}}}$
end for

If one is interested in the posterior mean and its covariance rather than the posterior mode for the position estimate then a Gibbs sampler [17] instead of the EM algorithm can be applied. However, the GS is significantly slower than the EM algorithm, because it requires significantly more samples (called iterations in the EM). In order to achieve the same positioning accuracy as the EM method within the test framework of [13] we had to use $T_0 = 500$ burn-in and $T = 1\,000$ retained samples, which resulted in a 400 times higher computation time for the GS compared with the EM. While for fitting parameters of skew-t distributions, which usually is done offline, the running time is of secondary importance and confidence intervals for the estimated parameters are more relevant, for online positioning this increase in computation time and the high computational demand can be critical and prohibit real-time positioning.

B. Dilution of precision

The precision of a trilateration algorithm is highly affected by the location of the receiver and BSs, which is defined by the geometry matrix $\mathbf{H}_{k,1:d}$ in (4). The position dilution of precision (PDOP) is a metric to assess the geometry quality of a certain position determination, which is written in [22, p.149] as

$$\mathrm{PDOP} = \sqrt{\mathrm{trace}\left\{\left(\mathbf{H}_{k,1:d}^{\mathrm{T}}\mathbf{H}_{k,1:d}\right)^{-1}\right\}}. \qquad (14)$$

The PDOP measures the confidence level of the position determination. A good geometry can be considered for PDOP values below 5, being excellent below 2.

IV. Results

This section assesses the performance of the EM and dGN algorithms in LTE realistic conditions. For this purpose, experimental results of these statistical trilateration techniques are obtained at the laboratory, by emulating real LTE signals and using an LTE software receiver. Standard channel models, i.e., EPA and ETU, are used to characterize the multipath environments for sake of reproducibility.

A. Explanation of the testbed

The experimental results obtained in this paper are based on the ranging measurements computed in an LTE testbed of the European Navigation Laboratory (ENL) at the European Space Agency (ESTEC, The Netherlands). As it is shown in Fig. 1, this testbed is divided in three parts:

1) *RF signal emulation*: LTE downlink signals are emulated at a certain RF band by means of the two Spirent E2010S network emulators, which are able to generate signals for up to 4 BSs. The channel effects are generated with a Spirent VR5 HD spatial channel emulator. The power of the received signals relative to the location of the receiver and BSs are applied according to pre-computed values with an LTE network simulator, such as in [23] and in [24]. The standard multipath channel models are then added to each downlink signal.
2) *RF signal capture*: the RF emulated signal is down-converted to baseband with a software-defined radio (SDR), i.e., USRP N210 with DBSRX2 daughterboard, which is a reconfigurable RF front-end. Then, the LTE baseband signal is digitalized and transferred to a computer, which stores the real and imaginary part of the captured samples in a file. An active hydrogen maser is used as an external reference clock for the SDR by generating a very stable 10-MHz reference signal.
3) *Baseband post-processing*: the baseband samples are post-processed with an experimental LTE software receiver in MATLAB. As it is described in [25], the main modules of this LTE software receiver are the cell detection, acquisition, tracking and positioning.

The testbed is configured to transmit up to 4 BSs on the LTE band 20, corresponding to a carrier frequency of 816 MHz, with a bandwidth of 1.4 MHz. The emulated network

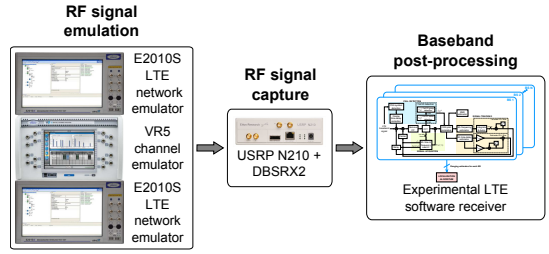


Fig. 1. LTE positioning testbed at ESA's Navigation Laboratory.

of BSs is tightly synchronized by the equipment, achieving a delay within one sample between the radio frames transmitted by different BSs. In addition, one BS is only used to generate LTE signals with a high SNR over an AWGN channel (that is, with LoS conditions) in order to aid the signal tracking at the receiver. The rest of BSs can be configured to emulate a specific scenario, by varying the multipath environment and the received power level. The delay corresponding to the signal time-of-flight between BS and receiver is not emulated in order to avoid synchronization and inter-cell interference errors at the receiver. The EPA5 and ETU70 models are here used to cover as many realistic scenarios as possible. The EPA5 model is aimed to represent open environments and pedestrians walking, whereas the ETU70 model is aimed to represent urban environments with high mobility.

The tests are conducted in two sequential phases: calibration and testing of the receiver. During the calibration phase, LoS conditions are emulated for every channel. This allows the receiver to estimate the network clock offsets. These clock offsets, which are typically within 10 ns, are obtained by averaging the time-delay estimates over the calibration period, i.e., 8 seconds in Section IV-B and 56 seconds in Section IV-D, approximately. During the testing phase, one BS remains in the same LoS conditions, while the specific multipath contribution is reestablished on the rest of emulated channels. This phase is used to test the trilateration algorithms proposed.

The LTE signals are captured by the USRP at a sampling frequency of 2 MHz with a gain of 31 dB. Then, the LTE baseband software receiver loads the data file of the signal capture and downsamples the signal to 1.92 MHz. The cell detection and acquisition is aided with the known cell ID of each BS. Given the tight network synchronization, signal tracking is entirely driven by one of the BSs that only transmits high-SNR signals over AWGN channel. The accurate time-delay and frequency estimates obtained using this BS provides tracking updates to the rest of BSs, which are transmitting LTE signals over multipath channels. Thus, one of the BSs is dedicated for accurate signal tracking, while the rest of BSs are used to obtain ranging measurements.

The tracking architecture of the LTE software receiver is based on a delay lock loop (DLL) and a frequency lock loop (FLL), as it is described in [26]. These tracking loops use the ML time-delay estimator and a usual frequency estimator [26]. The TDE range is bounded to one sampling period T_s, i.e., $[-0.5T_s, 0.5T_s]$, as in [7]. For a low signal bandwidth,

the resulting time-delay estimate can be considered similar to the one provided by threshold-based estimators [3], which are commonly used in the literature to mitigate multipath. Only 14 out of 40 CRS symbols per radio frame are considered, which are those aligned in the same subcarrier positions and with low inter-cell interference. These pilot symbols are used for TDE and frequency estimation, as well as for SNR estimation with the technique presented in [27]. Considering only the BS dedicated for signal tracking, its time-delay and frequency offset estimates are averaged every radio frame and filtered within the tracking loops. The sampling period of the loops T_{L} is then equal to 10 ms, and the noise bandwidth B_{L} is as tight as 0.5 Hz and 1 Hz for the DLL and FLL, respectively.

The ranging measurements are compensated by the network clock offsets computed during the calibration phase. These clock offsets include the time delay between the transmission time of the BSs, with respect to the local oscillator of the SDR. Thus, the compensated time-delay estimates are directly assigned to the ToA measurements, without any impact from tracking errors. In order to emulate the receiver location, the corresponding distance between receiver and BS is added to the ToA measurements. Thus, the performance of the statistical trilateration techniques is only affected by noise and multipath.

B. Fitted distributions for standard 3GPP channel models

In this section, we fit the parameters of a skew-t distribution to training data consisting of ranging errors from the two 3GPP channel models EPA5 and ETU70, with TDE ranges limited to the interval $[-0.5, 0.5]$ in T_s units, and nominal SNR around 30 dB. For comparison, we fit a normal distribution to the training data as well.

The training data consists of $n = 56\,000$ ranging errors for each of the two channel models, which are computed in separate tests. Two BSs are emulated for each test. The signal from the first BS is the most powerful and it is affected only by AWGN channel. This signal is used for tracking purposes in the coupled architecture described in the previous section. The second BS is received with a lower power than BS 1, and it is affected by EPA5 or ETU70 channel. For estimating the parameters of the skew-t distribution, the hierarchical model (7 – 8) and JAGS [28], which is a program that allows analyzing Bayesian hierarchical models by Gibbs samplers, are used with the second BS signal. We run the GS, as explained in Section II.C, twice with the initial parameter estimates $\xi_{(0)}$ set to the empirical mean and $\sigma^2_{(0)}$ set to the empirical variance of the errors; $\lambda_{(0)}$ is set to 0 and 1, and $\nu_{(0)} = 5$ for both runs. In both runs we use 200 "burn-in" and 500 retained samples. For each parameter, we apply a diffuse prior; only for ν a slightly informative prior, namely a uniform distribution over the interval $(2, 100)$ is used because $\nu > 2$ is required in (6).

The parameters of the normal distribution (mean μ and standard deviation σ_N) are fitted by maximum likelihood, which does not require any initial parameter estimates.

Table II contains the estimate statistics for the parameters of both fitted normal and skew-t distribution. Column *Mean* shows the mean estimate, and columns *5%-ile* and *95%-ile* show the 5 and 95 percentiles of each parameter.

TABLE II
Parameter statistics of normal and skew-t distributions fitted to ranging error data from EPA5 and ETU70 channel models. Values for restricted data are given in brackets.

Channel	Distr.	Para.	5%-ile	Mean	95%-ile
EPA5	normal	μ	15.7981	15.9552	16.1124
		σ_N	22.4975	22.6081	22.7198
	skew-t	ξ	15.7396	15.9815	16.1877
		σ	9.2301	9.2301	9.2998
		λ	-0.0257	-0.0008	0.0282
		ν	2.0001	2.0031	2.0089
ETU70	normal	μ	59.3554	59.7261	60.0968
			(52.9446)	(53.2947)	(53.5961)
		σ_N	53.0718	53.3325	53.6467
			(48.5168)	(48.7630)	(49.0120)
	skew-t	ξ	81.3757	86.6476	90.7608
			(88.6414)	(92.6285)	(94.4021)
		σ	48.8991	51.8804	54.5982
			(49.3767)	(53.4221)	(55.262)
		λ	-0.8541	-0.7162	-0.5484
			(-1.5437)	(-1.4413)	(-1.216)
		ν	7.0478	8.1329	9.3402
			(6.0753)	(7.3586)	(8.257)

Due to the large amount of training data the parameter estimates yielded by the applied methods are very certain for all parameters; the differences between the values in *Mean*, *5%-ile* and *95%-ile* are small.

For the EPA5 channel model, the ranging errors are unskewed, which causes similar values for the estimates of the normal distribution's mean μ and the skew-t distribution's location ξ. However, the pdf of the two fitted distributions differ significantly (see Fig. 2). While the normal distribution (black dashed line) is unable to capture the normalized error histogram's heavy tails and the concentration of probability mass around its peak, the skew-t distribution (blue line) captures the shape almost perfectly. This behavior could be expected due to the small estimate for the dof ν.

The normalized histogram of ranging errors from the ETU70 channel model also lacks significant skewness (see Fig. 3). In addition, the errors are more evenly spread than for EPA5. Thus, it is not surprising that the pdf's of both fitted distributions captures the histogram's shape well. Although the (unskewed) t-distribution approaches the normal distribution only as ν approaches infinity, already for values around 20 there is hardly any visible difference between a t-distribution and a normal distribution with same mean and scale. This and the small skewness, explains that both normal and skew-t fits are looking similar for the ETU70 channel.

In the histogram for the ETU70 channel model we notice a huge amount of errors in the right-most bin. This effect is expected because the estimator works within a small range, i.e., within one sampling period T_s. Therefore, any error out of the estimation boundaries will appear at the boundary value. The cause for error beyond the estimation boundaries is multipath. If we fit the distributions to the measurements without considering the ranging errors equal to the boundary

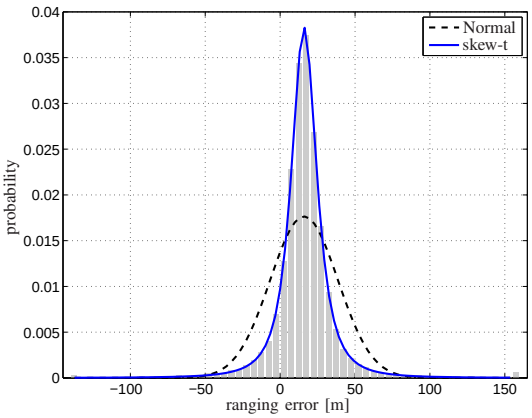


Fig. 2. Histogram for the EPA5 channel model with fitted Normal distribution and skew-t distribution.

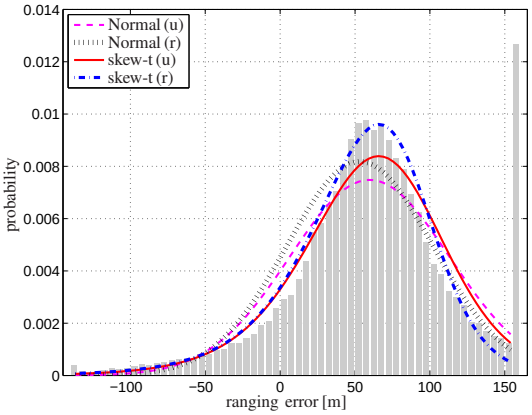


Fig. 3. Histogram for the ETU70 channel model with fitted Normal distribution and skew-t distribution to all ranging errors (labeled (u)) and fitted to all data except the ranging errors in the right-most bin (labeled (r)).

values, i.e., those ranging errors included in the right-most bin of the histogram, we obtain slightly different parameter estimates (see values in brackets in Table II). The pdf of the skew-t (blue dash-dot line) captures now the peak of the histogram much better, while the improvement for the normal distribution (black dotted line) is minor. This means, that the Gibbs sampler and the skew-t distribution are unable to filter out the outlier ranging errors produced by multipath. One reason for using the skew-t distribution was to reduce the impact of outliers. However, due to the large amount of those errors, here this distribution fails to do so, which was expected.

C. Numerical experiment

In this subsection we test the positioning with the EM algorithm for the EPA5 and ETU70 channel models with TDE range $[-0.5\,T_s, 0.5\,T_s]$, under the premise that ranging errors are skew-t distributed with the parameters given in column *Mean* of Table II. We compare the EM algorithm with the standalone dGN, which assumes normal distributed errors.

TABLE III
Positioning error quantiles for simulation experiment with additive skew-t measurement errors with the distribution's parameters taken from column *Mean* of Table II.

Channel	Data	Method	50%-ile	67%-ile	95%-ile
EPA5	all	dGN	9.11	11.14	20.23
		EM	6.24	8.03	16.81
ETU70	all	dGN	34.37	43.67	78.80
		EM	27.90	37.25	75.04
ETU70	restricted	dGN	38.55	43.88	65.44
		EM	19.90	26.75	56.35

The aim of the numerical experiment is to discover how both algorithms perform under perfect conditions in which measurement errors in training data and testing data are identically distributed. We will use the results in the next subsection to analyze both algorithms' performances when applied to emulated data, where the assumption of identical error distribution in general does not hold.

For this test, we use a similar setup as in [13]. In this section $\underline{\mathbf{x}}$ is a two-dimensional position and we assume it has the prior distribution

$$\mathbf{x} \sim \mathrm{MVN}\left(\mathbf{m}, \mathbf{P}\right) = \mathrm{MVN}\left(\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right], 1\,000\,\mathbf{I}_{2\times 2}\right), \qquad (15)$$

where $\mathrm{MVN}(\cdot,\cdot)$ denotes a bivariate Gaussian distribution with given mean and covariance matrix. Four BSs are located at the corners of a 1 000-by-1 000 square centered at $\mathbf{m}_0$, resulting in an excellent geometry with PDOP values below 2.

For each channel model, 1 000 receiver positions are drawn from the prior distribution (15). Using (3) for computing the true distance $h_k(\mathbf{x})$ between the receiver position and the BS location, $K = 12$ independent distance measurements (three to each BS) are drawn from $\underline{y}_k|(\underline{\mathbf{x}} = \mathbf{x}) \sim \mathrm{ST}\left(\xi + h_k(\mathbf{x}), \sigma^2, \lambda, \nu\right)$.

The hyperparameters used in the EM by the hierarchical version (13) of the measurement model are initialized as $\tau_k \leftarrow 1$ and $t_k \leftarrow -\xi_{\text{Median}}/\delta_{\lambda_{\text{Median}}}$, which ensures that the dGN finds the minimizer of the likelihood for Gaussian noise. For the EM algorithm 4 iterations are performed, and in each M-step 4 iterations of the dGN algorithm are performed. The dGN used for comparison also runs 4 iterations. In both the EM and the dGN algorithms the number of repetitions to find a suitable scaling factor α is limited to 5 [14].

Table III presents the error statistics for both algorithms, distinguishing for the ETU70 model between cases *all* (using parameters of skew-t fitted to all ranging errors in previous subsection) and *restricted* (using parameters given in brackets in Table II). The *50%-ile*, *67%-ile* and *95%-ile* are the quantiles of all two-dimensional positioning errors, which are defined as the Euclidean distance (cf. (3)) between the true position $\mathbf{x}$ and the position estimate $\hat{\mathbf{x}}$. The EM algorithm has always approximately five times higher computation time than stand-alone dGN, as explained in [13].

For all three cases the EM outperforms the dGN in each of the three error statistics. But rates of improvement differ

significantly. When using the parameters fitted to all available ranging errors the EM gives larger error reductions for the EPA5 model than for the ETU70 model. This behavior was expected, because for the EPA5 histogram (Fig. 2) the skew-t approaches the real pdf much better than the normal fit. For the ETU70 histograms (see Fig. 3), the differences are not that significant. However, the largest error reductions are achieved for the case where we use the parameters fitted to the restricted data, where we omitted the errors in the right-most bin of the ETU70's ranging error histogram.

It is important to note that we neglect the influence of the SNR in this test. That is, SNR is arbitrarily high and there is only a multipath and geometrical impact on the receiver localization, hence without considering the received power relative to the position of the receiver with respect to each BS. In the following section, a more realistic scenario is considered, where the SNR varies depending on the receiver position.

D. Test with real-world emulated data

Real-world conditions are tested in this section by emulating an LTE macro-cell deployment with testbed described in Section IV-A. Most of the possible receiver locations are tested by considering the received power at each point and EPA5 or ETU70 channel models. The performance and computational complexity of the EM and dGN algorithms are then assessed.

The test is based on an LTE network deployment over a 2 km-by-2 km area with seven BSs. This is the typical cell layout specified in [29] with three-sectorial macro-cells (i.e. 3 dB-beamwidth corresponding to 65 degrees), located in a hexagonal grid with inter-site distance of 750 meters (the arrangement of BSs can be seen in Fig. 6). The PDOP values of this deployment are between 1 and 2, resulting in an excellent geometry for positioning. The receiver locations are defined in a grid with step width of 50 meter, resulting in 1 681 grid points. The received signal power from each BS is computed as in [24], considering the standard network parameters in [29]. Since only ranging measurements from the three most powerful BSs are considered for each grid point, without lost of realism, only three BSs are emulated over EPA5 or ETU70 multipath channels for positioning purposes, while using one extra BS over AWGN channel for tracking purposes.

The LTE software receiver provides 280 measurements in a single grid point, because of 14 measurements per radio frame and 20 radio frames per grid point. Determining one position estimate uses four ranging measurements per BS, meaning that a total of 12 measurements are used in the EM or dGN algorithm. Given the 280 measurements per BS, we are able to compute 70 position estimates per grid point. The average of these estimates is then used to obtain the positioning error in that grid point.

For the EM algorithm, we assume ranging measurements being distributed as

$$y_k|\mathbf{x} \sim \mathrm{ST}\left(\xi_{\mathrm{Mean}} + h_k(\mathbf{x}), \sigma^2_{\mathrm{Mean}}, \lambda_{\mathrm{Mean}}, \nu_{\mathrm{Mean}}\right), \qquad (16)$$

and for the dGN algorithm, we assume ranging measurements being distributed as

$$y_k|\mathbf{x} \sim \mathrm{N}\left(\mu_{\mathrm{Mean}} + h_k(\mathbf{x}), \sigma^2_{N_{\mathrm{Mean}}}\right), \qquad (17)$$

where the parameter values are taken from column *Mean* in Table II.

We modify both methods such that the number of iterations in EM n_{EM} and dGN n_{dGN} are not fixed but rather using threshold values. If the position estimates of two consecutive iterations in either EM or dGN are closer than 1 meter to each other, the corresponding algorithm is terminated and the current estimate $\hat{\mathbf{x}}$ is returned. To avoid endless loops we use upper limits $n^{\max}_{\mathrm{EM}} = 50$ and $n^{\max}_{\mathrm{dGN}} = 50$; and similar to Section IV-B, we use a maximum of 5 iterations to find a suitable scaling factor α [14].

For both methods, we assume a multivariate-normal prior distribution for $\mathbf{x}$ with the mean being the center of the three most powerful BSs from which ranging measurements are available, i.e., the initial position estimate is considered at the barycenter of three most powerful BSs, and the covariance matrix is $\mathbf{P} = 1\,000^2\,\mathrm{m}^2\,\mathrm{I}_{2\times 2}$.

Positioning errors

Fig. 4 shows the empirical cdf's of the positioning errors in the tests. The results for the EPA5 scenario (left figure) are as expected, although the performance gain obtained using the EM rather than the dGN is smaller than anticipated based on the results in Section IV-C. For the ETU70 scenario, the error cdf produced by the dGN is better than that produced by the EM regardless of the use of parameters fitted to the whole or to the restricted training data. In addition, we notice that the EM performance depends clearly on the used parameters (red line vs. blue dash-dot line), while the dGN performance shows no significant dependence (magenta dashed line vs. black dotted line).

The main reason for the EM poor performance seems to be the varying SNR in the positioning test. The received power from each BS is different depending on the receiver location, being (in most of the cases) the signal of BS 1 the most powerful and the signal of BS 3 the most weak, as it is shown in the SNR histograms of Fig. 5. Thus, the noise affecting each ranging measurement may be more severe than the effect of multipath. As a result, the expected error distribution changes with respect to the prior distribution for the ranging measurement of each BS, decreasing the positioning performance of the statistical trilateration technique. Fig. 5 shows also the histogram of the average SNR per three BSs and per radio frame, i.e., every 10 ms or every 14 measurements, for the EPA5 and ETU70 scenarios. These histograms are compared with the SNR histogram of the training data used in Section IV-B. As it can be seen, the SNR of BS 2 and 3 are lower than those used in the training scenario, which we used for fitting the parameters of normal and skew-t distribution. In addition, the SNR values for the EPA5 channel model are higher than for the ETU70 channel model, because of the ETU70 model more frequent fading events.

Effect of SNR

We thus now analyze the performance of both algorithms within three SNR regions: high, medium and low SNR average, which are defined below. In each of the 1 681 grid points, we compute the average of the SNR values for the three

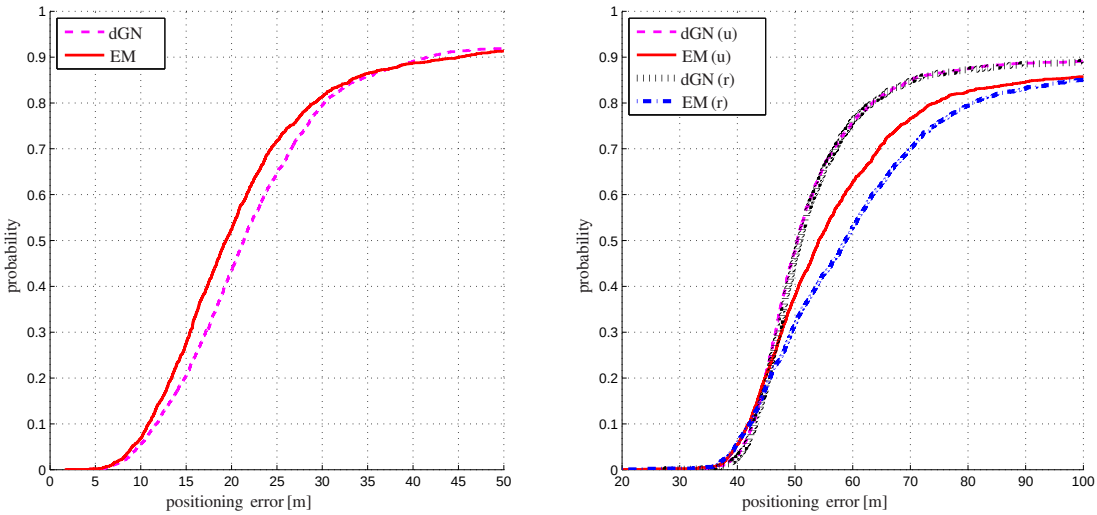


Fig. 4. Empirical cdf's of positioning errors for positioning tests of example topology. Left figure shows the cdf's for the EPA5 channel model and right figure shows the cdf's for the ETU70 channel model with parameter estimates fitted to both unrestricted (u) and restricted (r) training data.

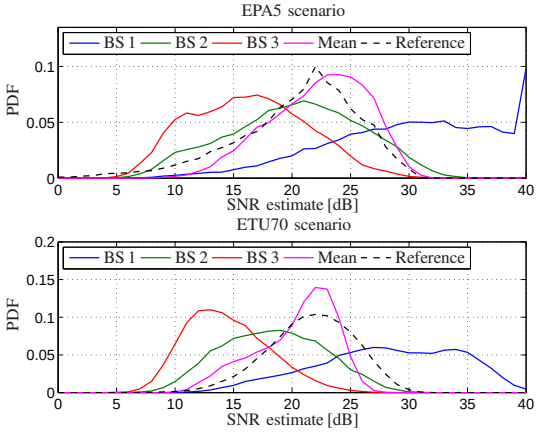


Fig. 5. SNR histograms for test data of EPA5 and ETU70 scenarios.

BSs observed in each point and label the point accordingly.

We label grid points as having a high SNR if their average SNR is at least as high as the 67% quantile of SNR values in the histogram for the specific channel, and as having a low SNR if their average is not larger than the 33% quantile; points with values in between are labeled as having medium SNR. Table IV shows for both scenarios the SNR intervals and the number of grid points with an average SNR inside the different regions plus their 67 and 95 percentile positioning errors.

The percentile errors show clearly that the positioning accuracy of both dGN and EM depend strongly on the average SNR, but for the EM the influence seems to be more severe. In the EPA5 test, the EM outperforms the dGN for scenarios with large average SNR and performs on a similar level for medium average SNR; but for low SNR averages the simpler dGN performs better.

A similar picture draw the errors in the two ETU70 tests, independent on which parameter estimates are used. However, for these tests the EM struggles already in the medium SNR range, which can be explained by the lower 33% and 67% quantiles of SNR values. Furthermore, the EM algorithm's rather good performance for lower SNR values for EPA5, compared with ETU 70, can be explained by the slower variation of the EPA5 channel. Therefore, there are more samples with almost the same multipath contribution, thus the parameter "refining" process, which is done by the EM, can be achieved with a short number of measurements.

The analysis supports our believe that for low SNR the EM algorithm is unable to correctly refine the channel parameters given the high SNR training data. It is more vulnerable to differences in SNR values between calibration (for fitting parameters of distributions) and positioning range measurements than the dGN. This vulnerability is the second price to pay, besides increased computation time, for the accuracy gain that we can get by using the EM rather than the dGN. By having similar SNR ranges in both fitting and positioning phase, we should be able to achieve better positioning accuracy with the EM than with the dGN.

An additional aspect we should still discuss are the large errors in both dGN and EM estimates for low SNR. Let us consider Fig. 6 in order to get a better understanding under which circumstances both algorithms fail to provide reasonable position estimates. It shows the EM position errors

TABLE IV
67 and 95 percentile positioning errors for real-world data experiment for different SNR regions. Column *No.* gives the number of grid points for which the average SNR of the 3 observed BSs is inside the interval presented in column *avg. SNR*.

Channel	avg. SNR [dB]	No.	Method	67% err [m]	95% err [m]
EPA5	$[25.26, \infty)$	387	dGN	17.33	29.05
			EM	15.85	26.49
	$(17.74, 25.26)$	1 067	dGN	25.66	46.27
			EM	23.54	51.51
	$[0, 17.74]$	227	dGN	322.82	1 506.07
			EM	373.44	1 801.22
ETU70 (all)	$[22.70, \infty)$	479	dGN	49.39	61.86
			EM	47.54	67.97
	$(15.79, 22.7)$	996	dGN	55.95	196.23
			EM	65.13	420.76
	$[0, 15.79]$	206	dGN	928.98	1 504.80
			EM	1 247.02	1 758.81
ETU70 (restr.)	$[22.7, \infty)$	479	dGN	50.92	62.05
			EM	48.50	72.70
	$(15.79, 22.70)$	996	dGN	56.00	215.90
			EM	70.98	425.15
	$[0, 15.79]$	206	dGN	945.07	1 513.09
			EM	1 256.33	1 751.60

for all 1 681 grid points in the ETU70 scenario, which used parameter estimates fitted to the whole training data. We limit the displayed errors to a maximum of 100 meters. For the other scenarios and also for the dGN we obtain similar heat maps.

The most precise position estimates are obtained when $\mathbf{x}$ is near a BS. The EM algorithm furthermore yields precise estimates if $\mathbf{x}$ is surrounded by the three observed BSs; it yields poor position estimates with errors of more than 100 meters if the true position has a poor geometry, i.e. if the BSs do not surround $\mathbf{x}$ and the BS antennas are facing away from it. This occurs because the received power from two of the three BSs is relatively low, resulting in poor ranging measurements that significantly degrade the position accuracy.

Complexity

As mentioned above, the EM requires more computations than the dGN. Thus, let us look at the number of iterations used in both dGN and EM algorithms. In [13], it was stated that those numbers specify the computational requirements of the two algorithms. Table V shows the mean as well as the 5 and 95 percentile values of the average number (over the 70 positioning attempts) of iterations used in a single grid point. For our threshold values of 1 meter for both dGN and EM, the EM and the standalone dGN use more than the 4 iterations that we used in [13], while for the dGN that is applied inside the EM, in general, less than 3 iterations are necessary. Thus, in our tests the increase in computational demand when switching from the dGN to the EM is on a similar level as shown in [13], i.e. approximately five times higher.

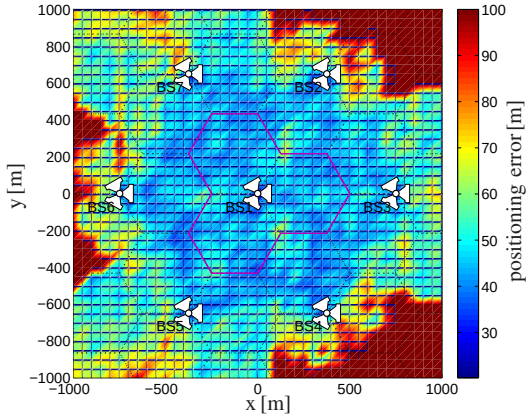


Fig. 6. Mean position error under ETU70 channel using EM algorithm, for which parameters were fitted to unrestricted data. Errors are limited to 100 meters. 65-degree sectors for each of the three antennas in each BS are denoted by triangles.

TABLE V
Average number of iterations used by dGN, EM and dGN used inside the EM algorithm (dGN-EM) in the test with realistic emulated LTE data.

Channel	Data	Method	5%-ile	Mean	95%-ile
EPA5	all	dGN	3.00	4.67	8.51
		EM	3.87	5.75	9.53
		dGN-EM	2.00	2.31	3.03
ETU70	all	dGN	3.11	5.15	8.63
		EM	5.80	7.23	10.72
		dGN-EM	2.22	2.71	3.66
	restricted	dGN	3.11	5.14	8.56
		EM	6.21	7.59	10.89
		dGN-EM	2.23	2.71	3.63

Influence of measurement quantity

Finally, we want to summarize our results on the influence of number of measurements used for positioning. The accuracy of both dGN and EM improves if more than 4 measurements from each of the 3 BSs are used, and worsens if less than 4 measurements are used. The shapes of the positioning error cdf's are similar to the ones in Fig. 4. For the ETU70 scenario, both in the restricted and unrestricted fitting approaches, the percentage of cases in which the EM performs better than the dGN increases slightly when less measurements are available. This can be explained by the fact that for smaller number of measurements it gets more complicated to filter out outlier measurements. Assuming skew-t distributed ranging errors simplifies this task compared with assuming a normal distribution.

V. Conclusions

In this paper we provide an extensive real-world application of the Expectation Maximization (EM) algorithm that was proposed in [13] for solving the trilateration problem for scenarios with skew-t distributed measurement errors. We

show how to fit the parameters of a skew-t distribution to training data using a Gibbs sampler, and compare those fits with fitted normal distributions. Data from two 3GPP standard channel models is used.

The focus of the paper is on the EM algorithm's positioning performance and the identification of the factors that influence it. We check how the EM method performs under ideal conditions and compare it with the performance of a standard Nonlinear Least Squares (NLS) algorithm. In the error quantiles we see improvements of up to 39%. The EM method's better performance is due to using a more accurate statistical model.

However, when applied to real-world data from an emulated LTE network, we have realized that EM performance is degraded in the presence of SNR disparities between training and the actual measurements; and this happens for instance when the SNRs from different BSs noticeably differ. In cases where the SNR in training and test data are similar the EM outperforms the NLS; but when the SNR in the training data is significantly higher than in the test data, the EM accuracy deteriorates more than the NLS's, causing it to perform worse than the NLS. This can be explained by the fact that varying SNR changes the distribution of ranging errors, and the EM can refine the assumed error distribution only to a certain degree. Furthermore, in the tests the fitted skew-t distributions described the ranging error training data more precisely than the fitted normal distributions. Thus, the EM using these skew-t distributions naturally suffers more from a change in error distribution than the NLS using normal fits. The parameters of the fitted normal distributions will change less for a changed error distribution than the parameters of the fitted skew-t distributions, which can be seen, for example, in Fig. 3.

Future work should address this issue by fitting skew-t distributions to training data with different SNR values, and then use the fitted distribution whose SNR is closest to the SNR in the trilateration problem. In addition, it should be determined how exactly a changing SNR changes the ranging error distribution and how mathematical models for the parameters could be used to address these changes.

REFERENCES

[1] 3GPP TS 36.305, *E-UTRAN; Stage 2 functional specification of UE positioning in E-UTRAN*, Std., Rel. 12, 2014.

[2] 3GPP TS 36.211, *Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 8)*, 3rd Generation Partnership Project Std., V8.6.0, (2009-03).

[3] D. Dardari, C.-C. Chong, and M. Win, "Threshold-based time-of-arrival estimators in UWB dense multipath channels," *IEEE Transactions on Communications*, vol. 56, no. 8, pp. 1366–1378, August 2008.

[4] Z. Xiao, H. Wen, A. Markham, N. Trigoni, P. Blunsom, and J. Frolik, "Non-line-of-sight identification and mitigation using received signal strength," *IEEE Transactions on Wireless Communications*, vol. 14, no. 3, pp. 1689–1702, March 2015.

[5] H. Chen, G. Wang, Z. Wang, H. C. So, and H. V. Poor, "Non-line-of-sight node localization based on semi-definite programming in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 1, pp. 108–116, January 2012.

[6] W. Xu, M. Huang, C. Zhu, and A. Dammann, "Maximum likelihood TOA and OTDOA estimation with first arriving path detection for 3GPP LTE system," *Transactions on Emerging Telecommunications Technologies*, 2014.

[7] J. A. del Peral-Rosado, J. A. López-Salcedo, G. Seco-Granados, F. Zanier, and M. Crisci, "Joint maximum likelihood time-delay estimation for LTE positioning in multipath channels," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 33, p. 13, February 2014.

[8] X. Li and K. Pahlavan, "Super-resolution TOA estimation with diversity for indoor geolocation," *IEEE Transactions on Wireless Communications*, vol. 3, no. 1, pp. 224–234, January 2004.

[9] P. Müller, H. Wymeersch, and R. Piché, "UWB positioning with generalized Gaussian mixture filters," *IEEE Transactions on Mobile Computing*, vol. 13, no. 10, pp. 2406–2414, October 2014.

[10] M. Kok, J. D. Hola, and T. B. Schön, "Indoor positioning using ultrawideband and inertial measurements," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 4, pp. 1293–1303, April 2015.

[11] N. A. Alsindi, B. Alavi, and K. Pahlavan, "Measurement and modeling of ultrawideband TOA-based ranging in indoor multipath environments," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 3, pp. 1046–1058, March 2009.

[12] B. Alavi and K. Pahlavan, "Modeling of the distance error for indoor geolocation," in *2003 IEEE Wireless Communications and Networking Conference (WCNC 2003)*, vol. 1, March 2003, pp. 668–672.

[13] P. Müller and R. Piché, "Statistical trilateration with skew-t errors," in *Proc. 2015 International Conference on Localization and GNSS (ICL-GNSS)*, June 2015, pp. 1–6.

[14] R. Piché, "Estimation of model parameters," in *Mathematical Modeling with Multidisciplinary Applications*, X.-S. Yang, Ed. John Wiley and Sons, 2013, pp. 169–190.

[15] A. Azzalini, *The skew-normal and related families*. Cambridge University Press, 2014.

[16] T. I. Lin, J. C. Lee, and W. J. Hsieh, "Robust mixture modeling using the skew t distribution," *Statistics and Computing*, vol. 17, no. 2, pp. 81–92, June 2007.

[17] A. E. Gelfand and A. F. M. Smith, "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 398–409, 1990.

[18] 3GPP TS 36.101, *Evolved universal terrestrial radio access (E-UTRA); User equipment (UE) radio transmission and reception*, 3rd Generation Partnership Project Std., Rel. 9, V9.18.0, January 2014.

[19] 3GPP TR 36.873, *E-UTRA; Study on 3D channel model for LTE*, Std., Rel. 12, V12.1.0, March 2015.

[20] L. Xu and M. I. Jordan, "On convergence properties of the EM algorithm for Gaussian mixtures," *Neural Computation*, vol. 8, pp. 129–151, 1995.

[21] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[22] J. Sanz, J. M. Juan, and M. Hernández-Pajares, *GNSS Data Processing, Vol. I: Fundamentals and Algorithms*. Noordwijk, the Netherlands: ESA Communications, ESTEC TM-23/1, 2013. [Online]. Available: http://www.navipedia.net/GNSS_Book/ESA_GNSS-Book_TM-23_Vol_I.pdf

[23] C. Mehlführer, J. Colom-Ikuno, M. Šimko, S. Schwarz, M. Wrulich, and M. Rupp, "The Vienna LTE simulators–Enabling reproducibility in wireless communications research," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, pp. 1–14, 2011.

[24] J. A. del Peral-Rosado, J. A. López-Salcedo, G. Seco-Granados, F. Zanier, and M. Crisci, "Achievable localization performance accuracy of the positioning reference signal of 3GPP LTE," in *Proc. International Conference on Localization and GNSS (ICL-GNSS)*, June 2012, pp. 1–6.

[25] J. A. del Peral-Rosado, J. M. Parro-Jiménez, J. A. López-Salcedo, G. Seco-Granados, P. Crosta, F. Zanier, and M. Crisci, "Comparative results analysis on positioning with real LTE signals and low-cost hardware platforms," in *Proc. 7th ESA Workshop on Satellite Navigation User Equipment Technologies (NAVITEC)*, December 2014, pp. 1–8.

[26] J. del Peral-Rosado, J. Lopez-Salcedo, G. Seco-Granados, P. Crosta, F. Zanier, and M. Crisci, "Downlink synchronization of lte base stations for opportunistic toa positioning," in *Proc. International Conference on Localization and GNSS (ICL-GNSS)*, June 2015, pp. 1–6.

[27] M. Zivkovic and R. Mathar, "An improved preamble-based snr estimation algorithm for OFDM systems," in *Proc. IEEE 21st International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, Sept 2010, pp. 172–176.

[28] September 2013. [Online]. Available: http://mcmc-jags.sourceforge.net

[29] 3GPP TR 36.942, *Evolved universal terrestrial radio access (E-UTRA); Radio frequency (RF) system scenarios*, Std., Rel. 9, V9.3.0, July 2012.

Tampereen teknillinen yliopisto
PL 527
33101 Tampere

Tampere University of Technology
P.O.B. 527
FI-33101 Tampere, Finland

ISBN 978-952-15-3769-1
ISSN 1459-2045